



# Optimal Hyperparameters for AI-based Hip Fracture Detection

**Presented by**

Ms. Praechompoo Nilbanjong 665040046-8

**Advisor**

Assoc. Prof. Dr. Puripong Suthisopapan, Department of Electrical Engineering, Faculty of Engineering, Khon Kaen University

**Co-advisors**

Asst. Prof. Dr. Prin Twinprai, Department of Radiology, Faculty of Medicine, Khon Kaen University

Dr. Nattaphon Twinprai, Department of Orthopedics, Faculty of Medicine, Khon Kaen University

**Committee**

Assoc. Prof. Dr. Khanit Matra, Department of Electrical Engineering, Faculty of Engineering, Srinakharinwirot University  
Assoc. Prof. Dr. Anan Kruesubthaworn, Department of Electrical Engineering, Faculty of Engineering, Khon Kaen University  
Assoc. Prof. Dr. Pirat Khunkitti, Department of Electrical Engineering, Faculty of Engineering, Khon Kaen University

# Outlines

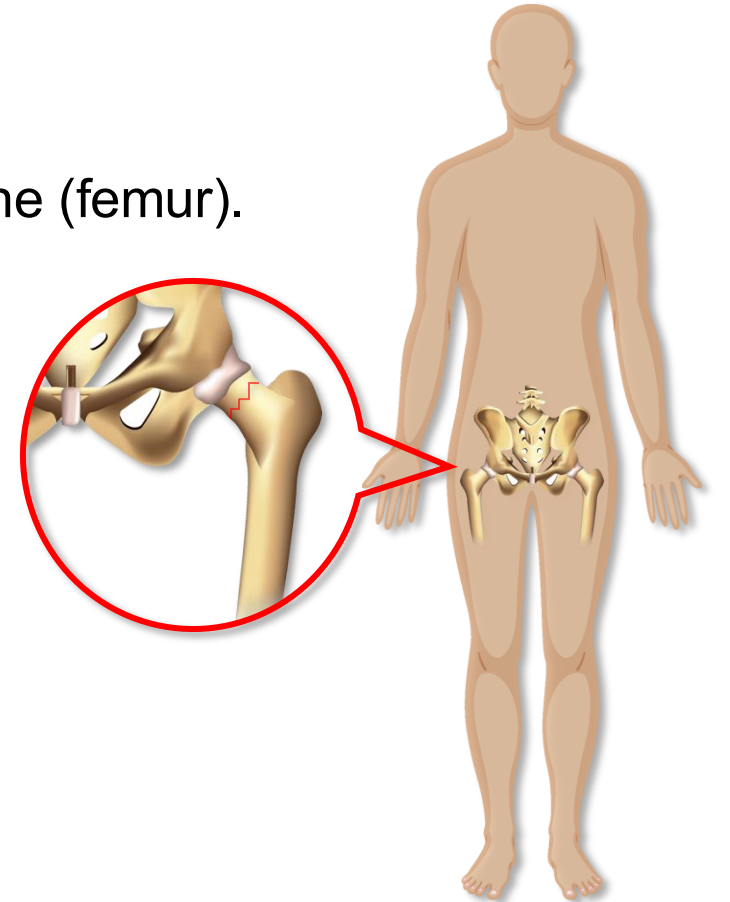
- 1 Introduction
- 2 Objectives and Scopes
- 3 Theoretical Background
- 4 Literatures Reviews
- 5 Methodology
- 6 Results
- 7 Limitations and Future Works

# Introduction

# Hip Fracture

## *What is hip fracture?*

- Hip fracture is a **break or crack** in the upper portion of thigh bone (femur).
- The region connects to the pelvis bone.
- It is **one of the most** prevalent **types of bone fractures**.



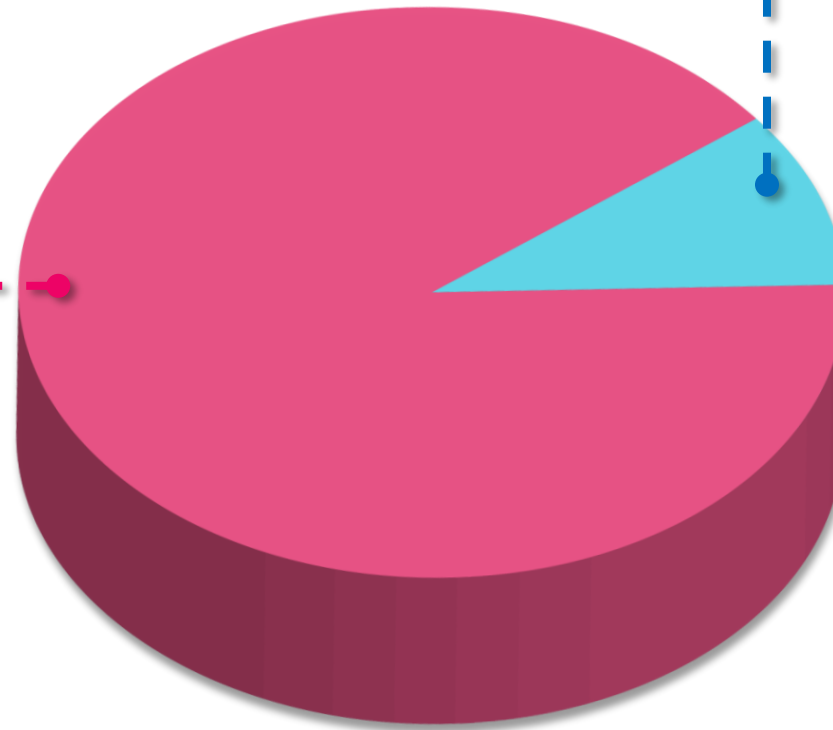
# Hip Fracture (cont.)

## Causes

**90 %**

physical factors:

- Accident
- Fall



**10 %**

- Physical inactivity
- Body anthropometrics
- Muscle weakness
- Medications and alcohol
- Chronic health condition
- Bone structure

# Hip Fracture (cont.)

## *Fact*

- Hip fracture can occur in individuals of **all genders and ages**.
- Most prevalent in older adults **around 60 years old**.
- According to estimates, **approximately 2.6 million per year by 2025**.

**“Hip fracture cases are expected to  
double by 2050”** 

# Hip Fracture (cont.)

## **Fact**

Hip fractures occur more frequently in **women than in men.**

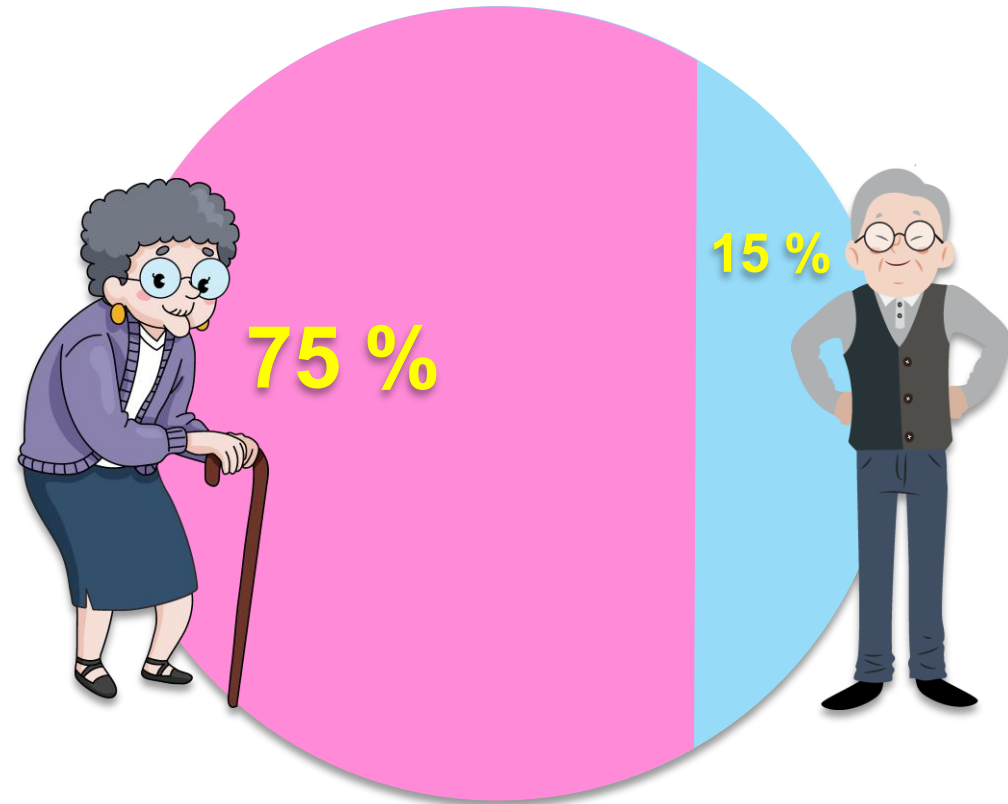
Postmenopausal women



Bone mineral density



Risk of fracture



**“ Many countries transition into an aging society ”**

# Hip Fracture Diagnosis

## *Fact*

The diagnosis of hip fracture is made through visual inspection of **medical imaging**.

X-ray image



CT scan



MRI image



# Hip Fracture Diagnosis (cont.)

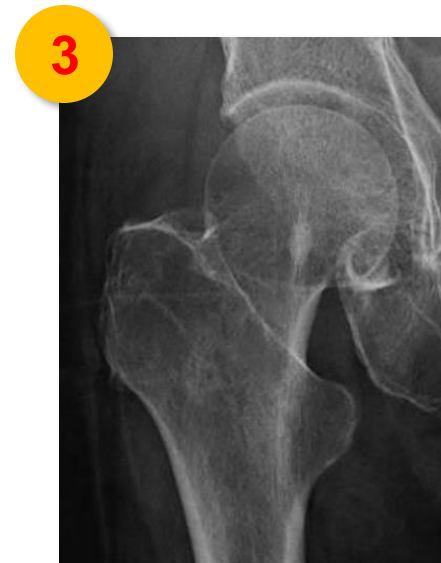
## ***Fact***

- Clinical recommend, **reparative surgery in 24 - 48 hours.**
- The 1-year mortality rate can be as high as 50% with patients without surgery.
- Many areas lack specialized physicians for diagnosis hip fracture.
- **Missed diagnoses as high as 14 - 30%.**

**“The rate of misdiagnosis is quite high!”**

# Hip Fracture Diagnosis (cont.)

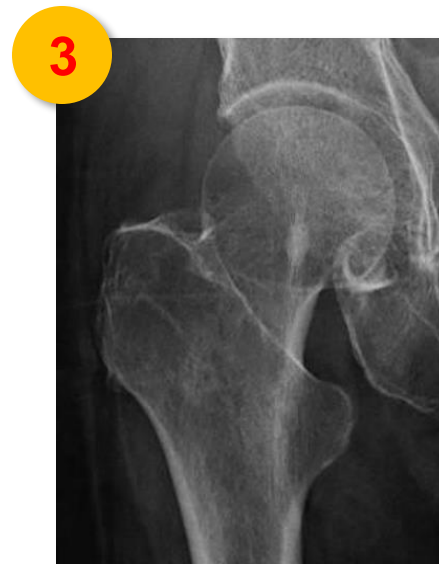
*Which choice is hip fracture?*



# Hip Fracture Diagnosis (cont.)

*Which choice is hip fracture?*

**It is difficult to diagnose with naked eyes!**



The answer is....

**NUMBER 2**

# Challenges in Hip Fracture Diagnosis

No fracture



Femoral neck fracture



Intertrochanteric fracture



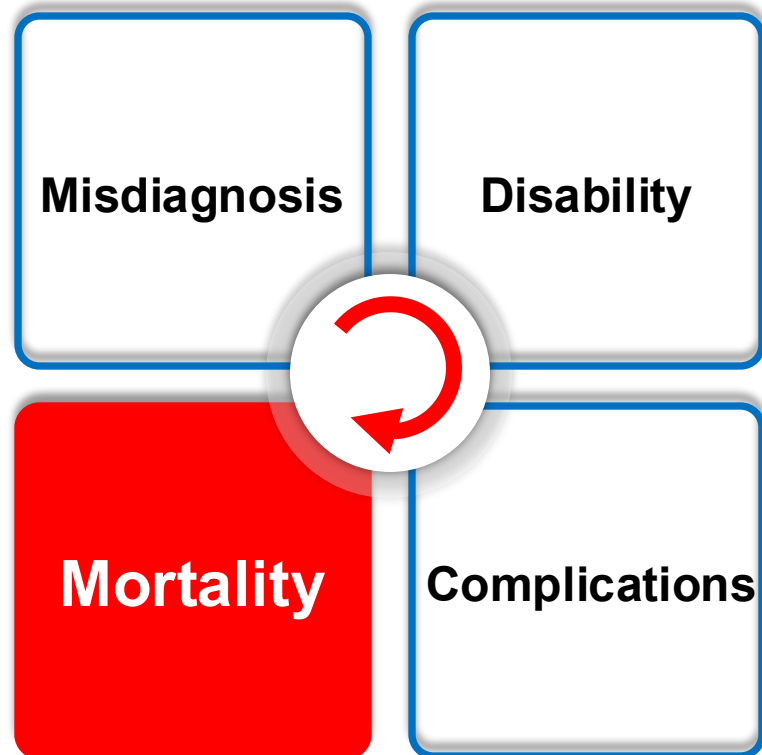
Subtrochanteric fracture



- **Complex anatomy** of physical characteristics.
- Limited diagnostic **accuracy and time**.
- **Lack of clinical experience**.

# Effects of Misdiagnosis

## *Fact*

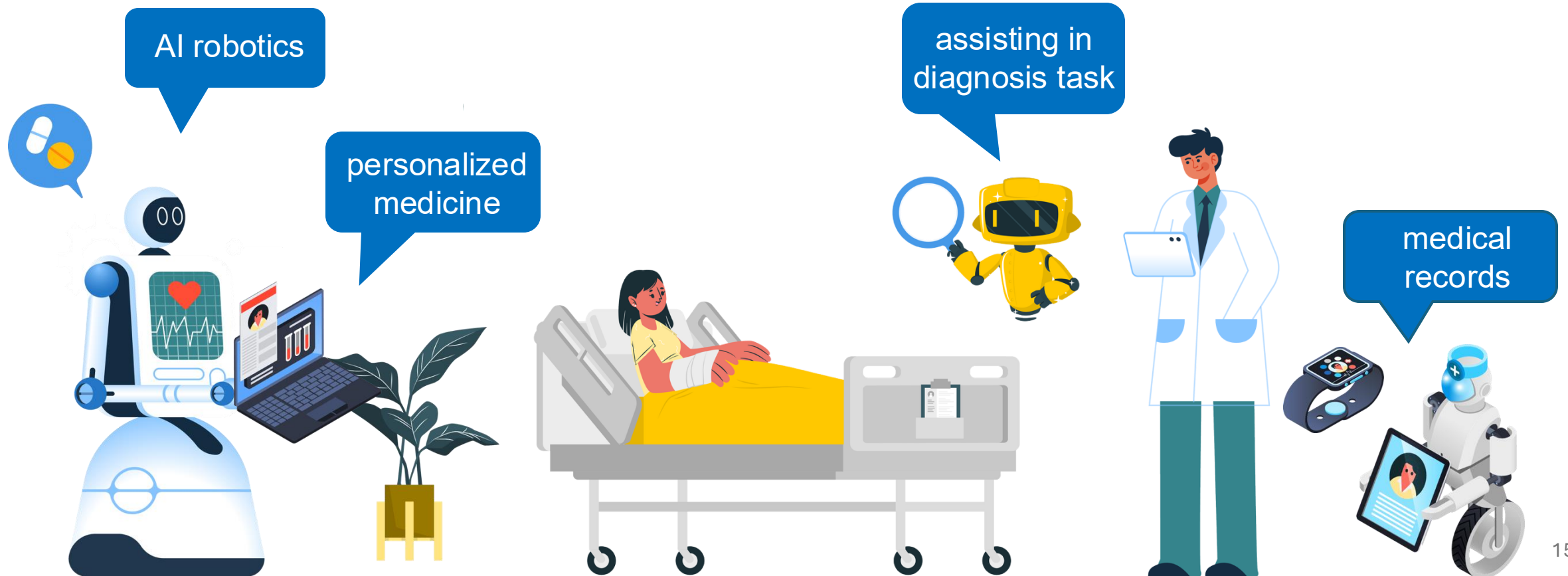


- Significantly **impacts patient outcomes**.
- **Up to 20% mortality within 1 year.**
- Approximately 30% of survivors **may be left with long-term disability**.



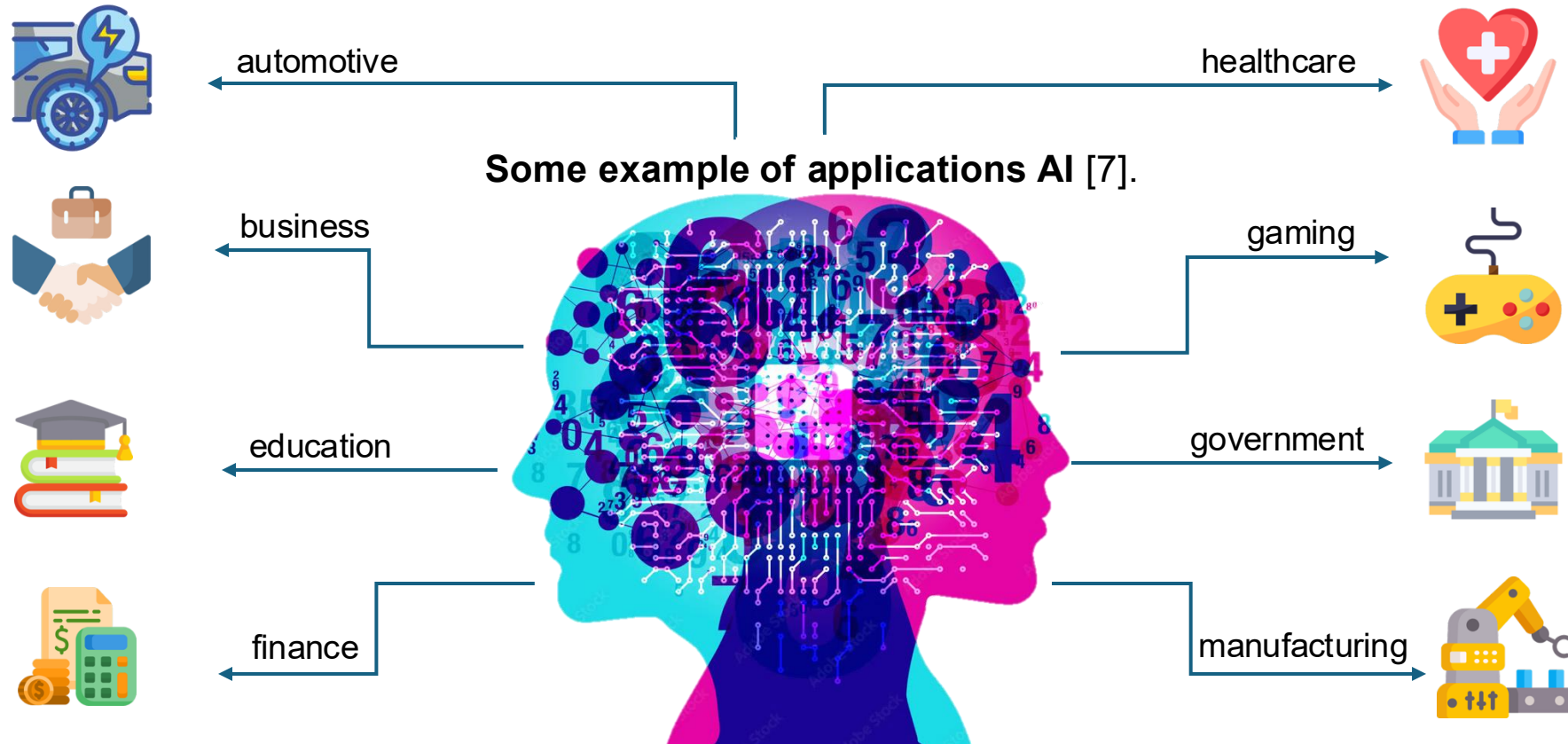
# AI Healthcare Application

## *Example of AI healthcare applications*



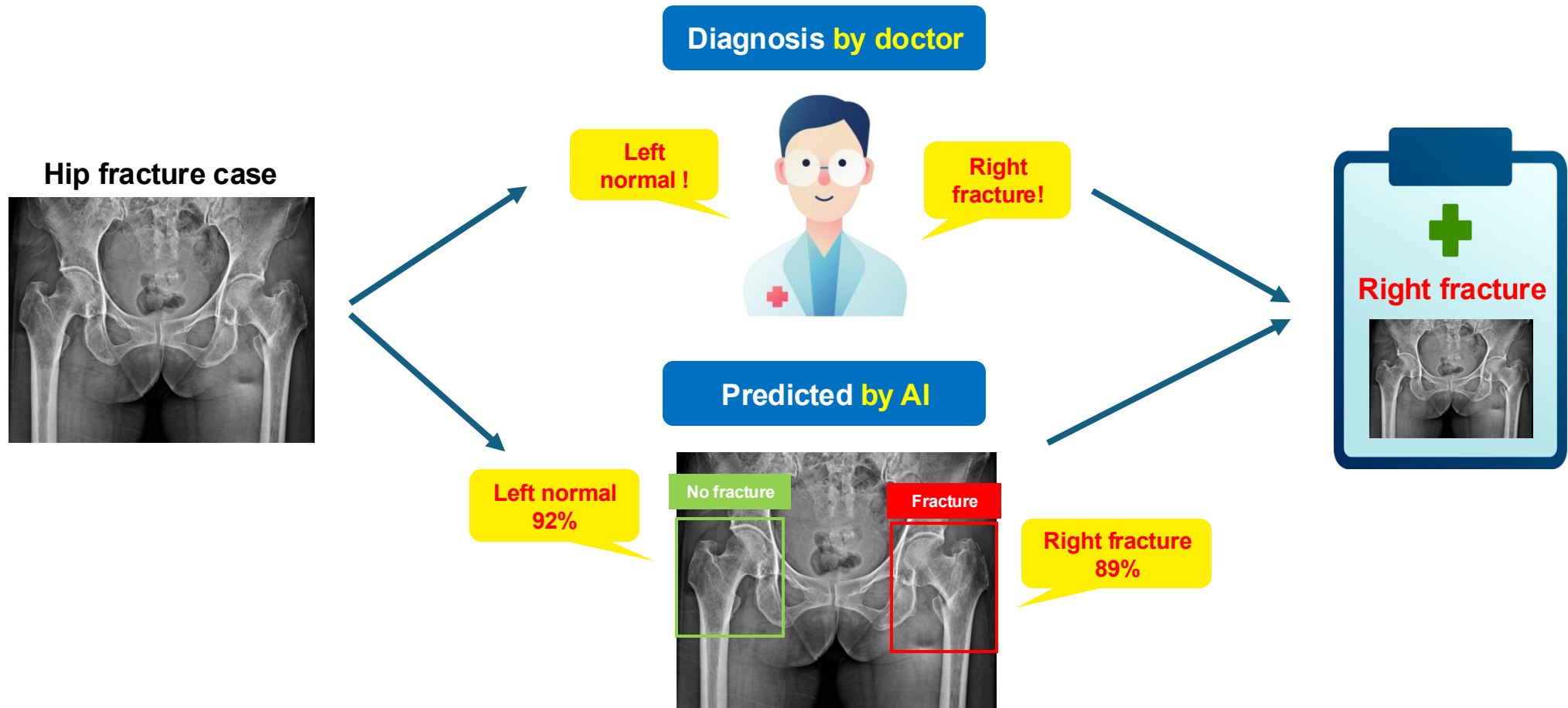
# What is AI?

- Artificial intelligence or AI.
- Learn and simulation of human behavior processes.



# AI-assisted Medical Diagnosis

## Example of AI in hip bone diagnosis



# AI-assisted Medical Diagnosis (cont.)

Year	AI in hip bone diagnosis	Techniques
2020	Automatic detection and functional subclassification of hip fractures using deep learning	DenseNet
2021	Deep learning for femoral neck fracture detection on pelvic radiographs	DCNN
2022	Machine learning in classification of hip fracture	CNNs
2022	Comparison of AI and Human Performance in Hip Fracture Detection	YOLOv4-tiny
2023	Radiographic detection of hip dysplasia using deep learning	YOLOv5
2023	Non-displaced femoral neck fracture detection on hip radiographs	ViT
2024	Deep learning model for bone fracture detection	YOLOv8 and VGG16

**“Small margin of diagnostic errors even a 1%,  
significant impact in clinical decision-making”**

# Motivation

Year	Optimal hyperparameter
2022	An improved hyperparameter optimization using evolutionary algorithms
2022	Optimizing the hyperparameter tuning of YOLOv5 for underwater detection
2023	Deterministic selection-based genetic algorithm for hyperparameter tuning
2023	Automatic image detection in YOLOv5 and compare analysis of YOLOv8
2023	Optimization of YOLOv7's hyperparameter for object detection
2023	Modeling 3D printer fault detection using improved YOLOv8 with hyperparameter optimization
2023	Impact of deep learning optimizers and hyperparameter tuning on the performance of bearing fault diagnosis

“ Optimal hyperparameters **improve the performance** of the AI models ”

# Objectives and Scopes

# Objectives

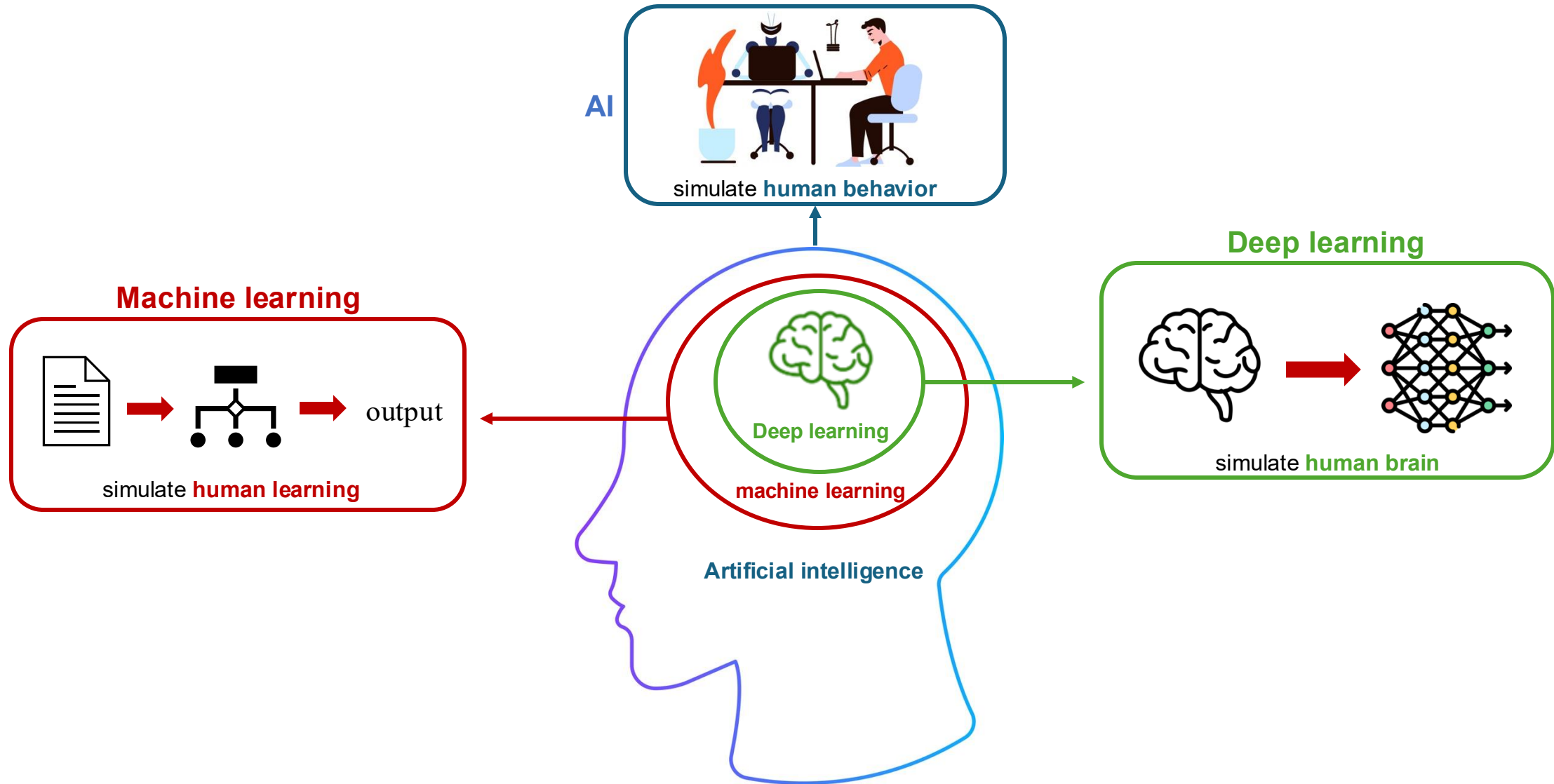
- 1 To apply the YOLO algorithm to develop models for hip bone X-ray images.
- 2 To determine the optimal hyperparameters for hip fracture detection based on the YOLO algorithm.

# Scope of the Study

- 1 Development of the detection model based on the YOLO algorithm.
- 2 Utilization of hyperparameter tuning based on a genetic algorithm.
- 3 Collection of 1,580 hip bone X-ray images from Srinagarind Hospital, Khon Kaen University.

# Theoretical Background

# Artificial Intelligence



# Machine Learning

**Machine learning** is machines learn relationships between inputs and outputs through data using mathematical processes.

## Supervised learning

Apple



Banana



The model learned from **labeled data** with **known output** during training.

## Unsupervised learning



The model **identifies patterns** or **structures** from unlabeled data.

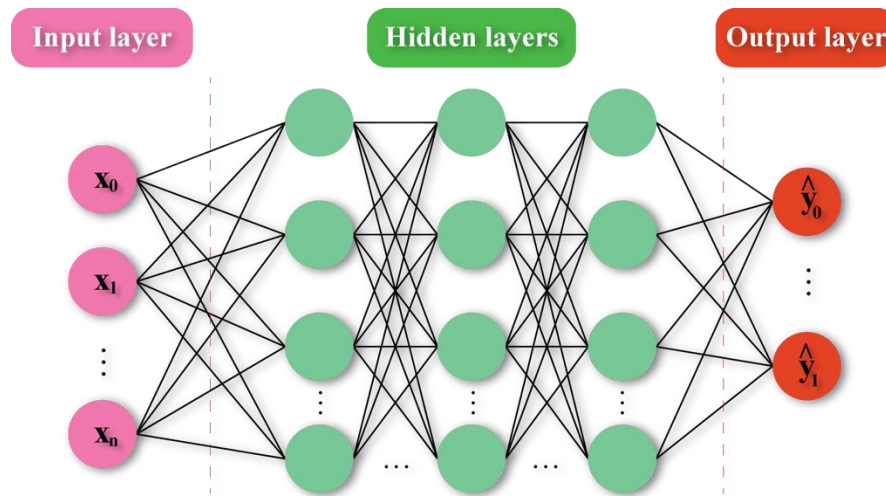
## Reinforcement learning



The model learns by **trial and error** through **feedback and rewards**.

# Deep Learning

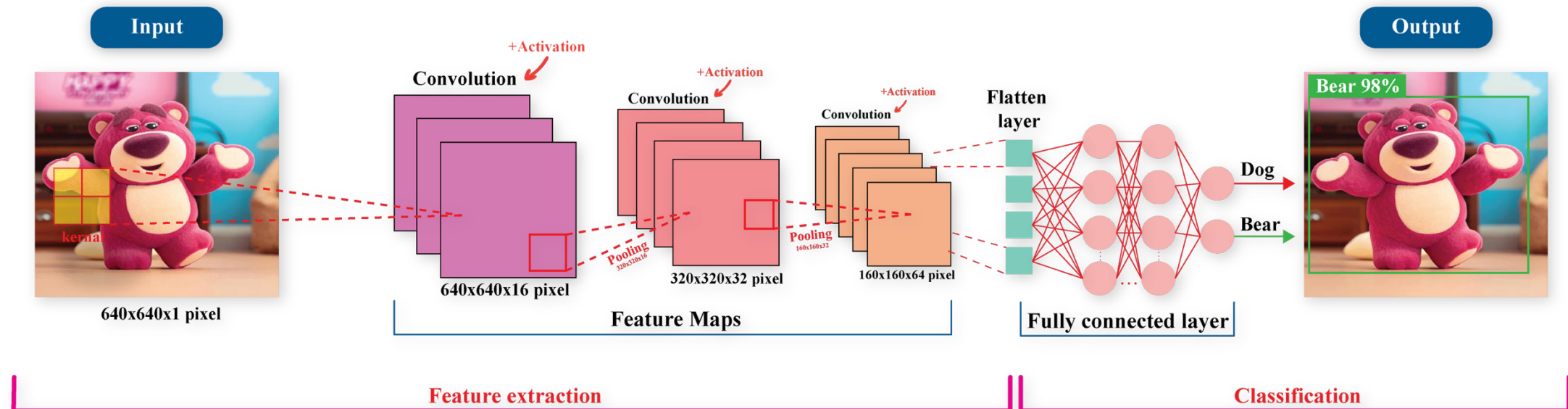
**Deep learning** is a learning methods attempting to model data **complex architectures**.



- **Input layer** is a first layer receives input data and each node represents a difference feature.
- **Hidden layer** is a layer connect between input and output layer.
- **Output layer** is final layer computes the final output of neural network.

# Convolutional Neural Networks

- Convolutional neural networks (CNNs) are a subset of ANNs in the field of deep learning.
- It is a specialized kind of neural network for **finding patterns of images**.



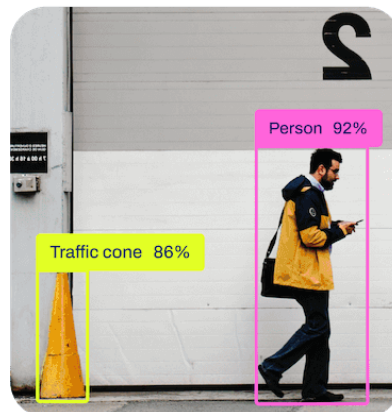
# YOLO

- You Only Look Once (YOLO) [12].
- **Object detection system.**
- Single forward propagation through a neural network (end-to-end).
- **High speed and high accuracy.**

Classify



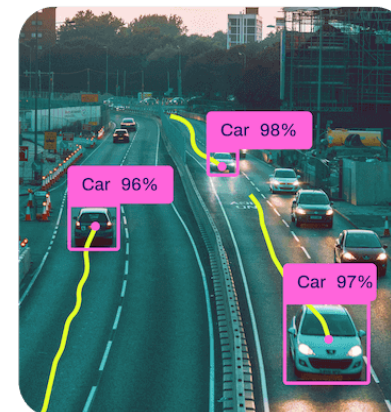
Detect



Segment



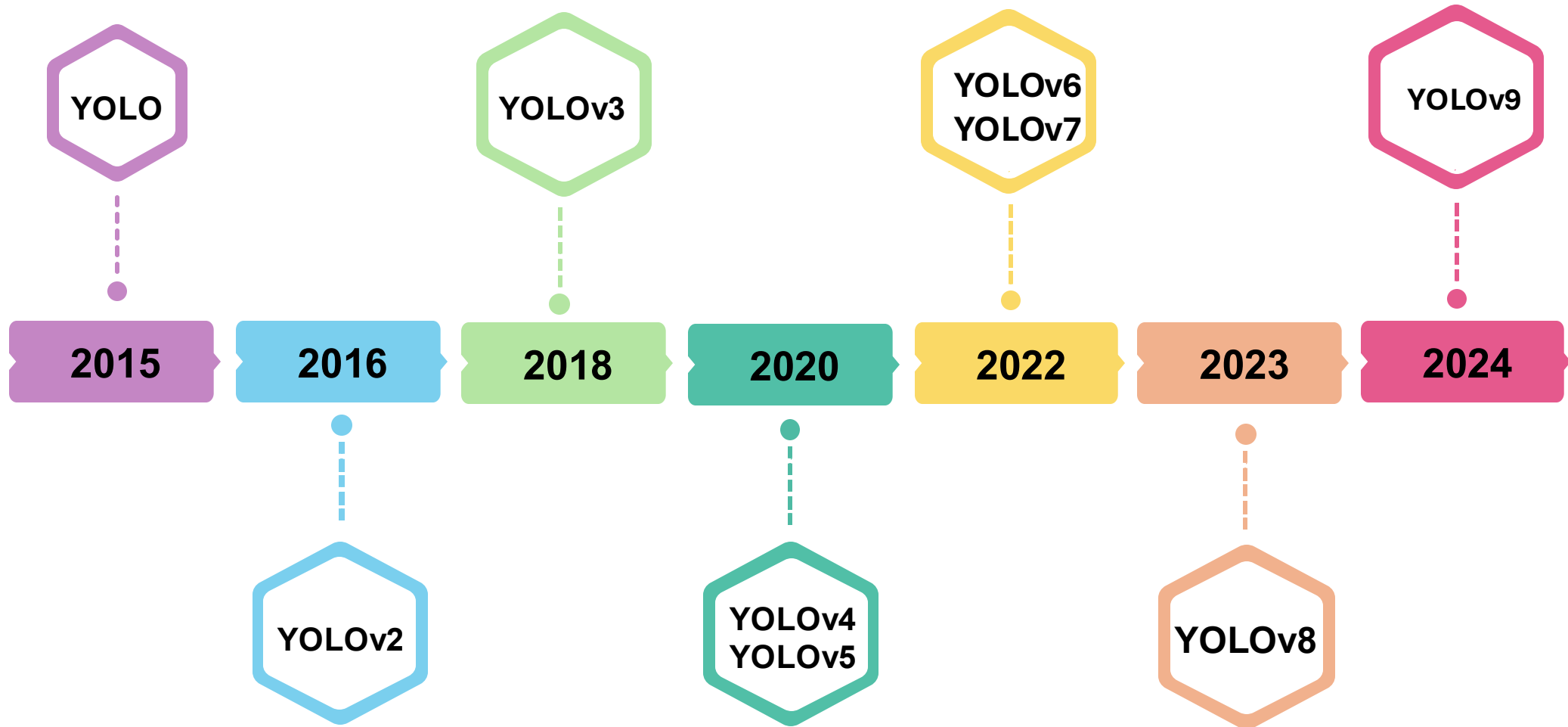
Track



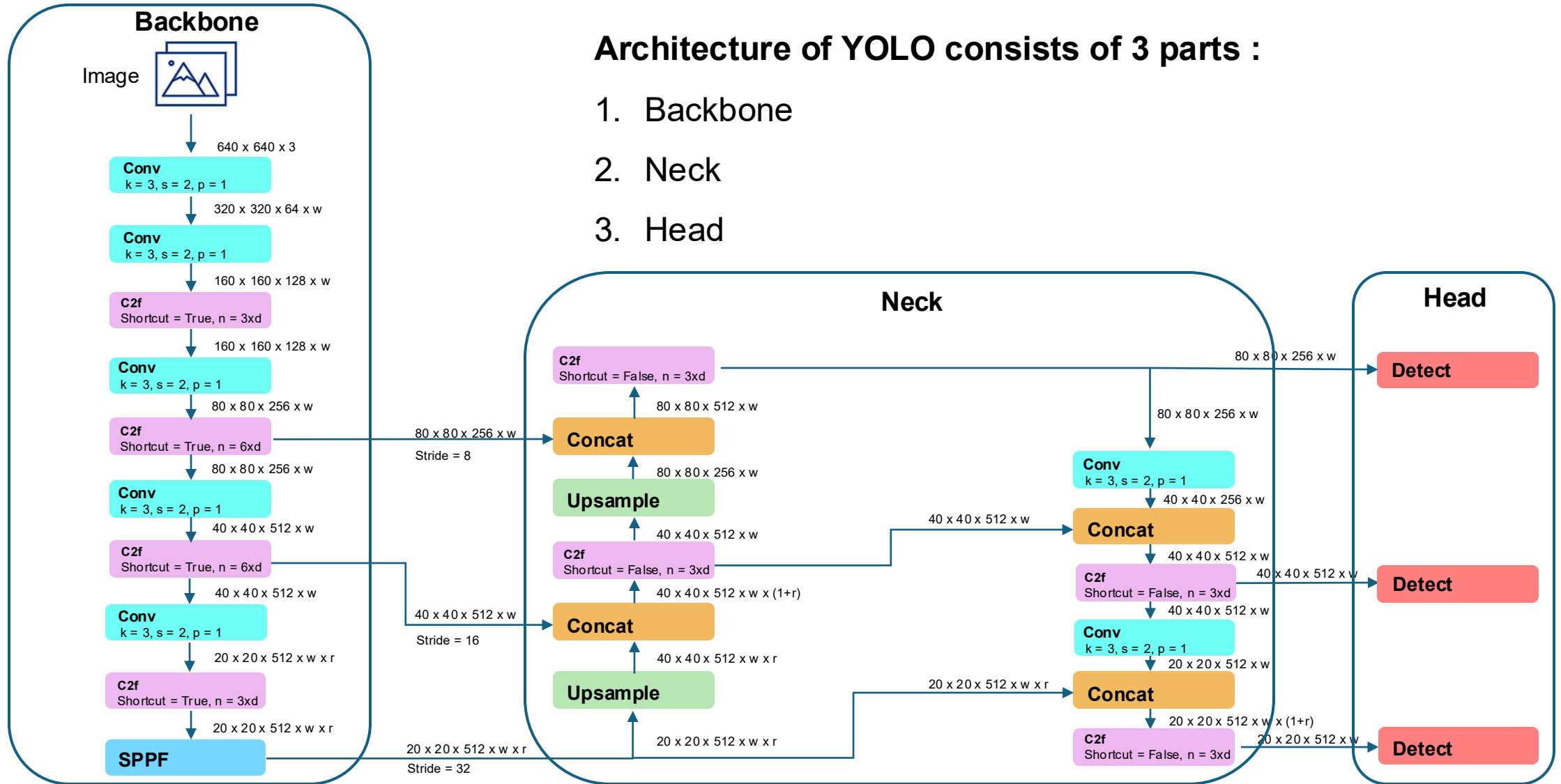
Pose



# YOLO History



# YOLO Architecture



Architecture of YOLO consists of 3 parts :

1. Backbone
2. Neck
3. Head

# Hyperparameter Tuning

In order to apply YOLO **for hip fracture detection**, must be jointly tuned.

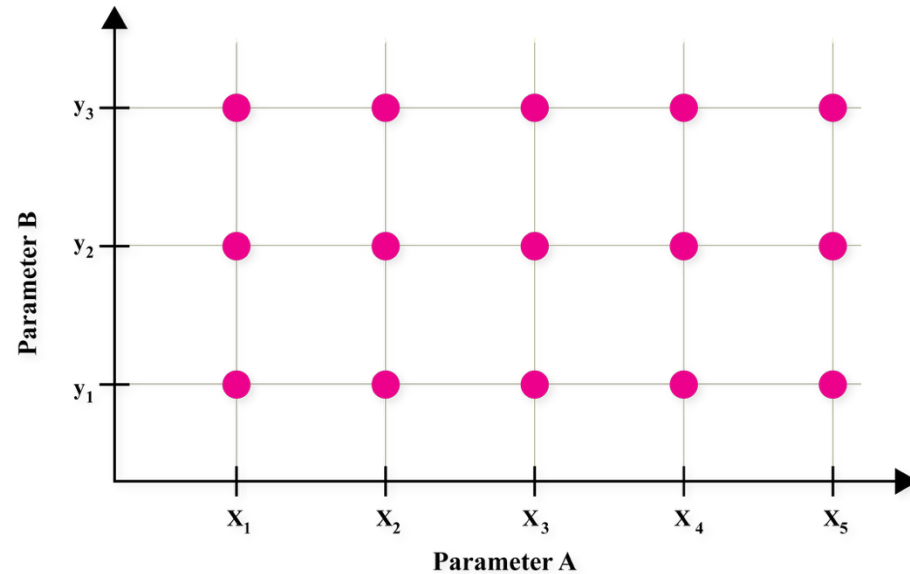
- Epochs
- Batch
- Learning rate
- Optimizer
- Momentum

and many more [9] ....



So, that the **optimal performance** can be achieved.

# Grid Search

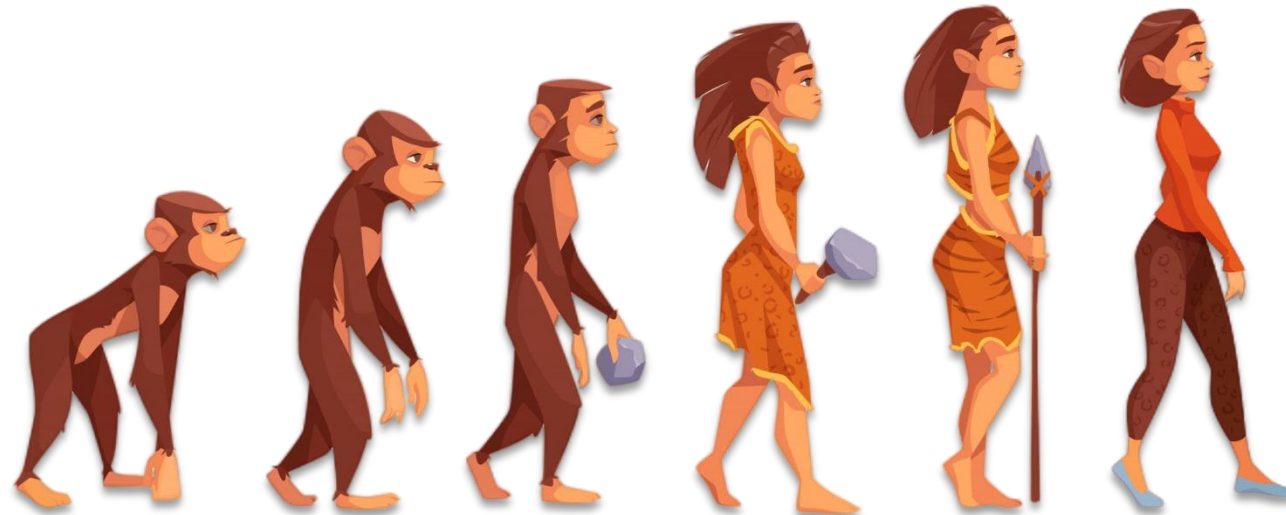


- **Grid search** is an optimization method.
- **Explore all possible combinations of hyperparameter** value within a defined ranges of values.

# Genetic Algorithm

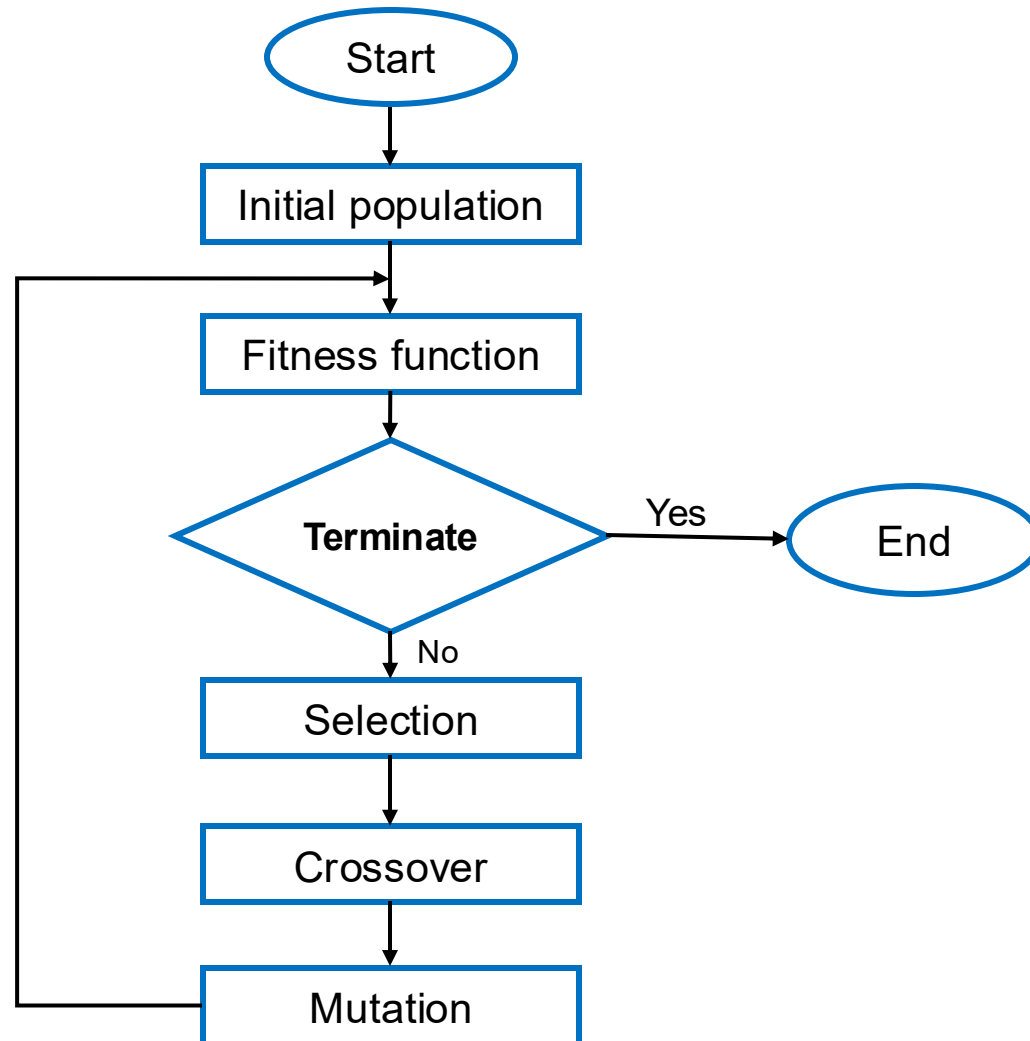
- Genetic algorithm or GA.
- Inspired by biological evolution technique.
- Use **nature selection process.**
- **Finding the best specific condition.**
- To solve optimization problem.

**Highly effective  
for solving complex problem**



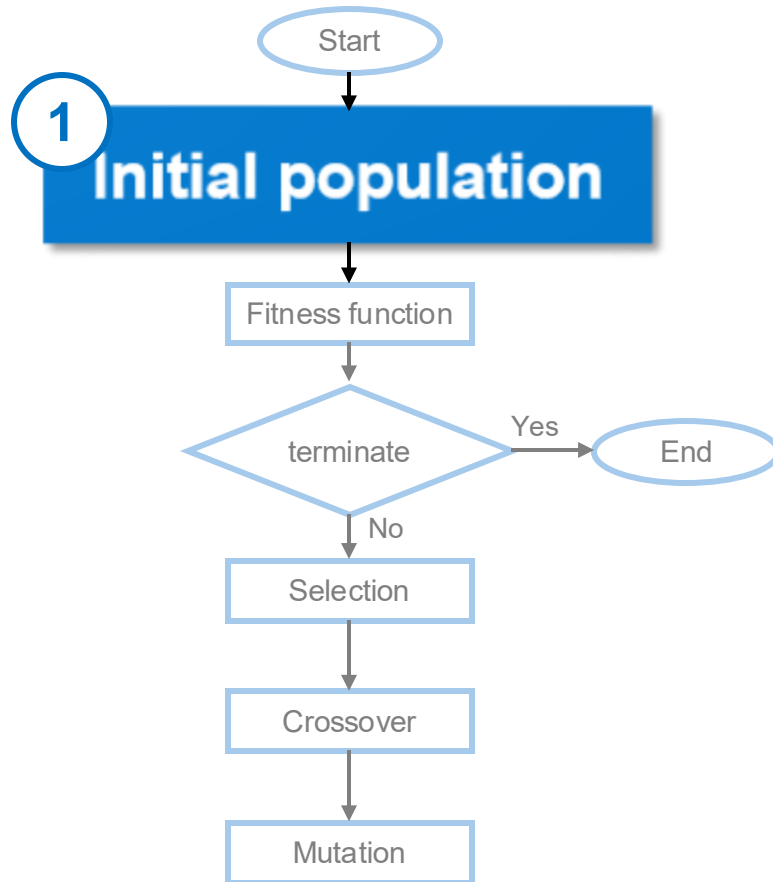
# Flow Chart of Genetic Algorithm

## *Definition*



# Initial Population

## Definition

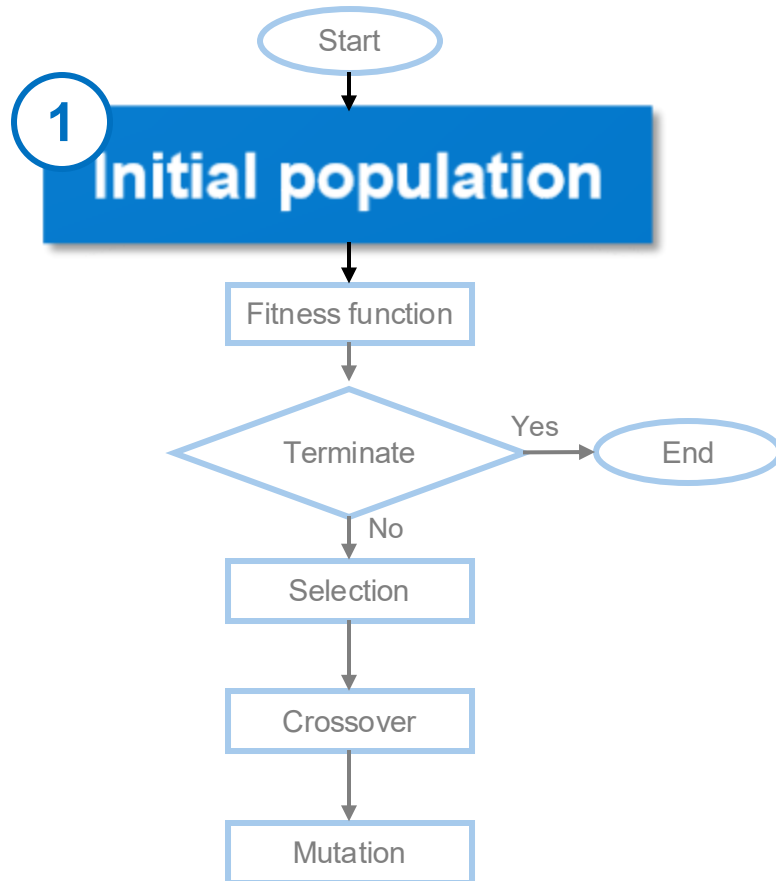


## Random initial population

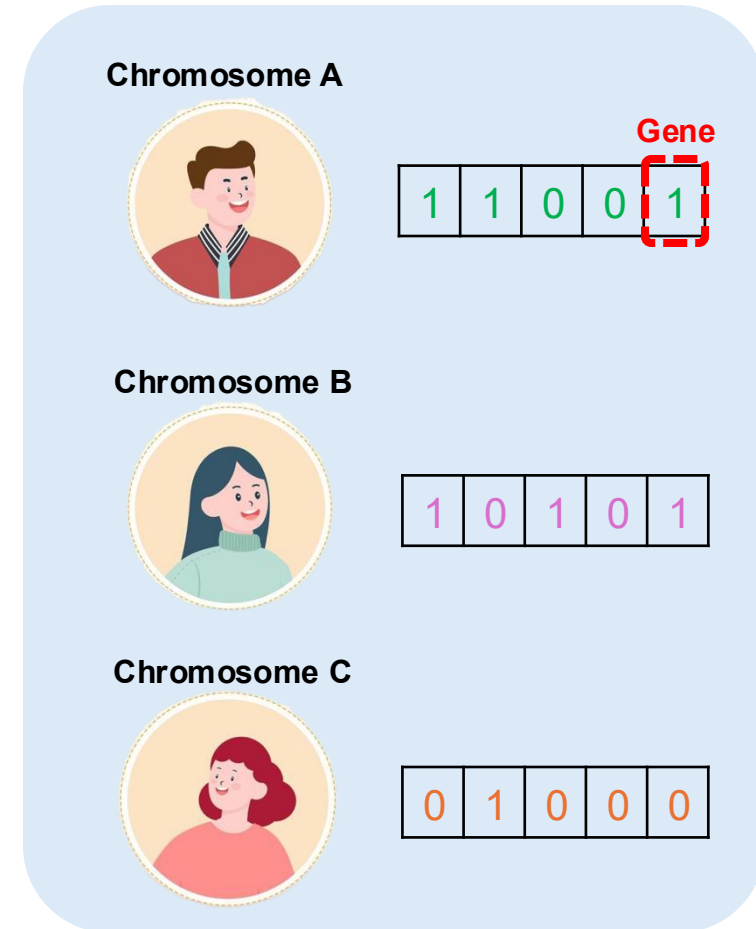


# Initial Population (cont.)

## Definition

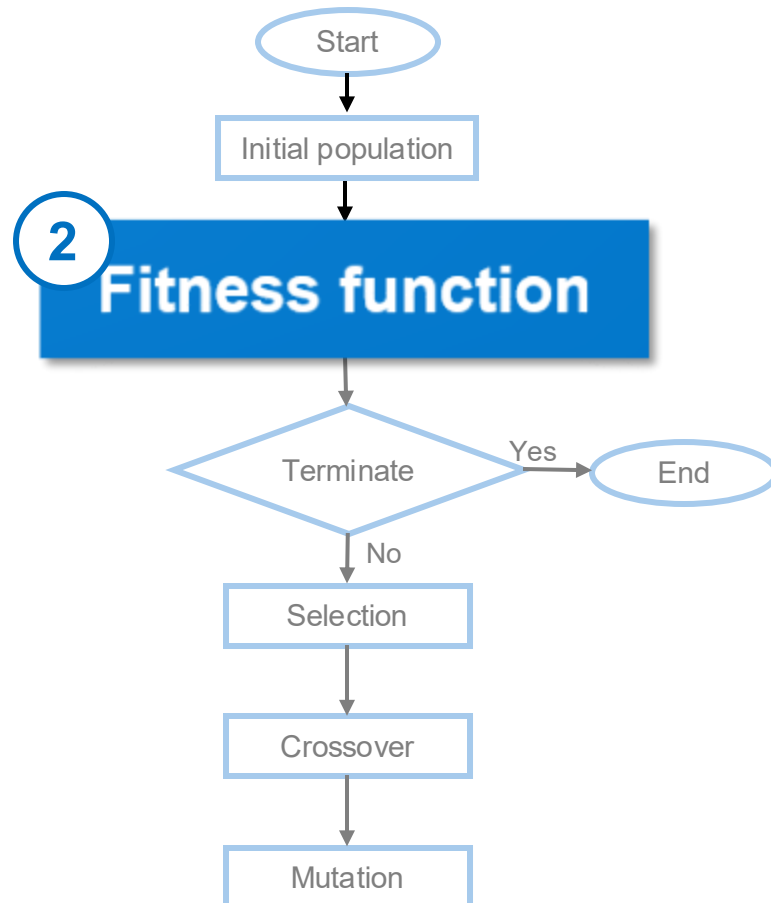


## Population



# Fitness Function

## Definition



“A fitness function **evaluates a solution** in relation to the optimal or **best solution.**”

# Fitness Function (cont.)

2

## Fitness function

EX

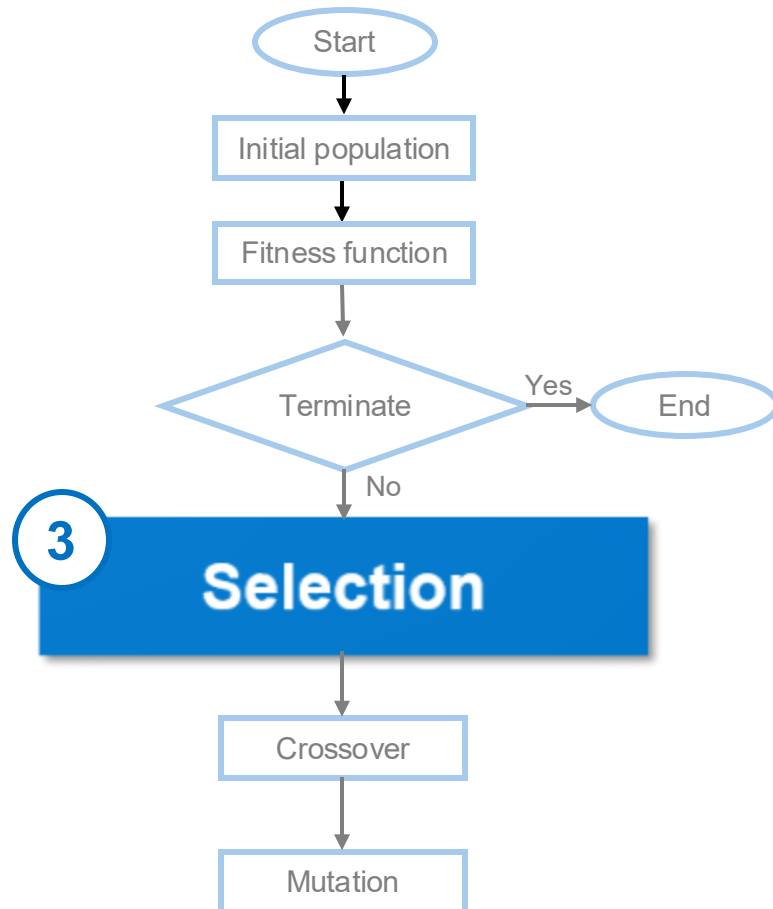
$$f(x) = x^2$$

➤ Given the minimum value is 0 and maximum value is 30.

Chromosome	population	x value	Fitness value
A	11001	25	625
B	10101	21	441
C	01000	8	64

# Selection

## Definition



Chromosome	population	x value	Fitness value
A	11001	25	625
B	10101	21	441
C	01000	8	64

**Chromosomes A and B** have the best fitness value.

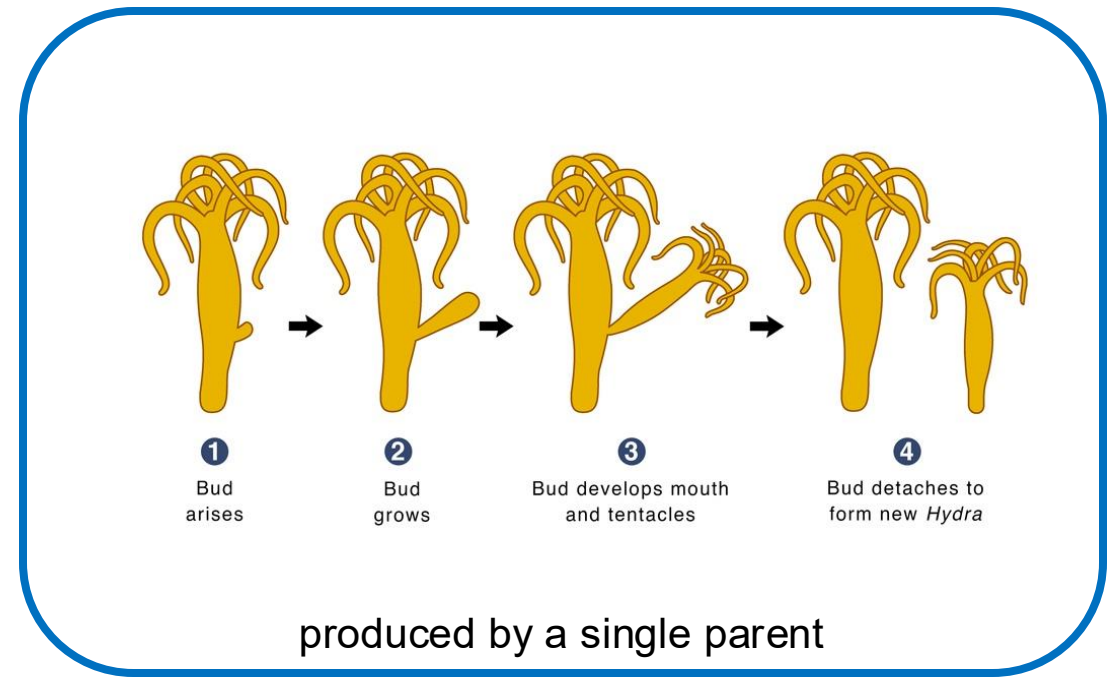
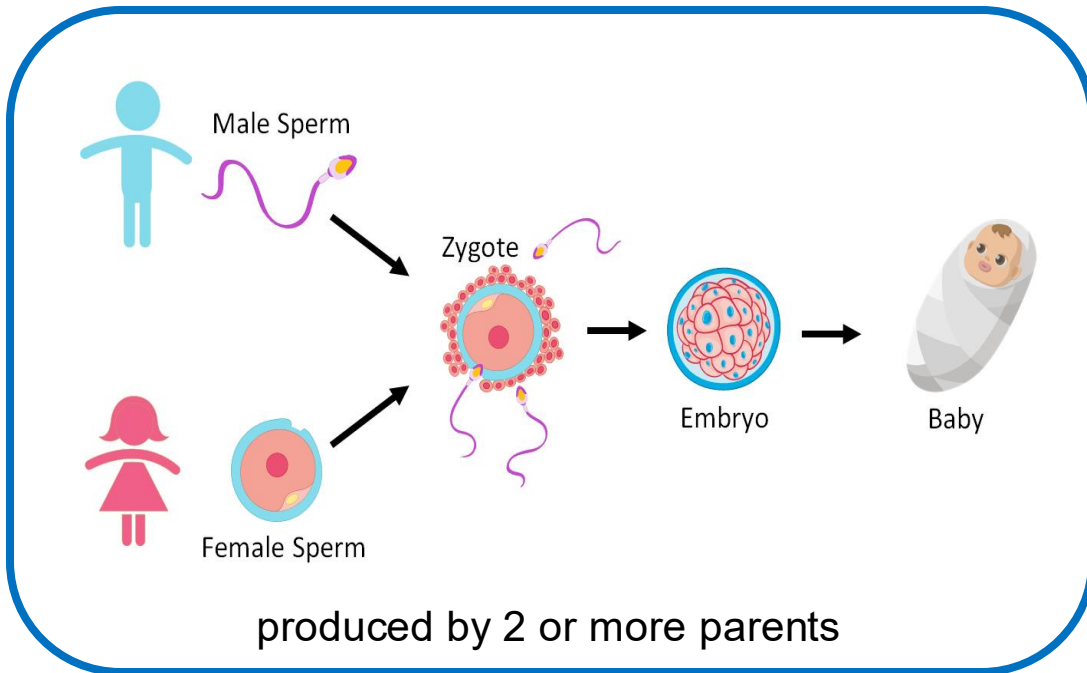
# Reproduction

## Definition

## Reproduction

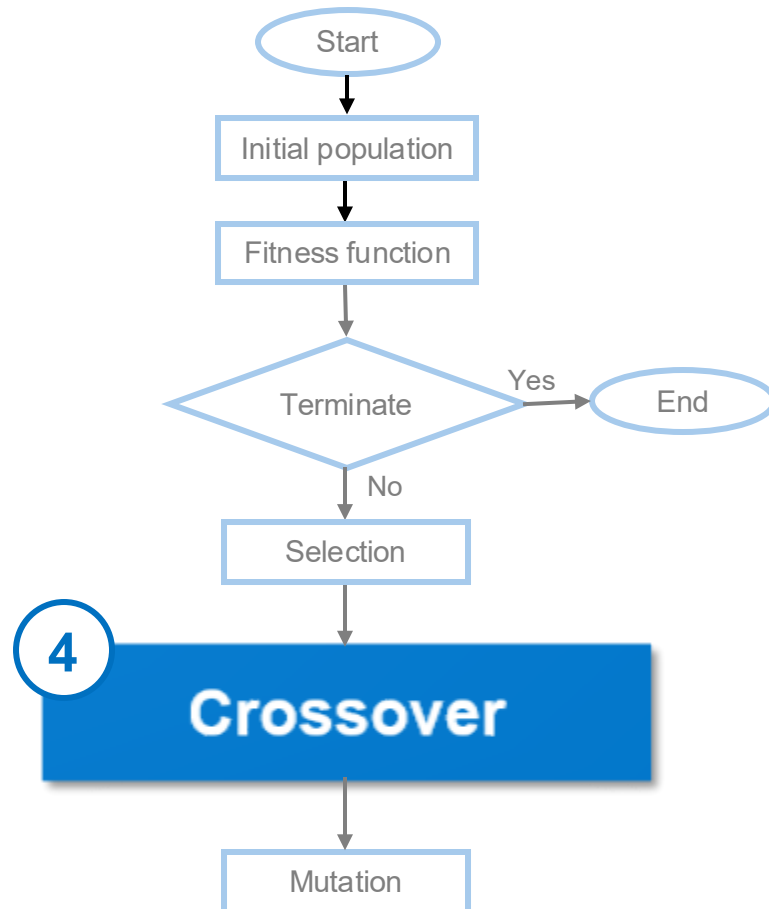
Sexual reproduction

Asexual reproduction

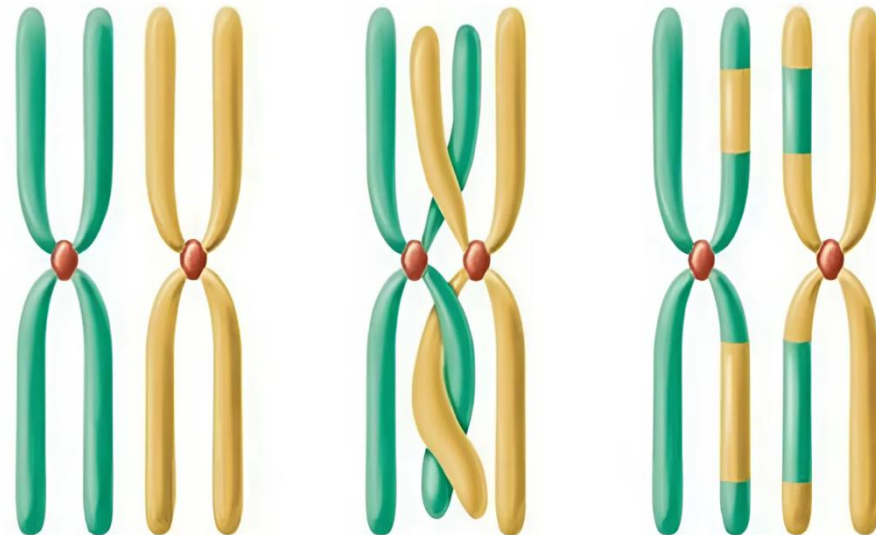


# Crossover

## Definition



“**Crossover** is sexual reproduction to process the **generating new offspring** through **between 2 or more parents.**”



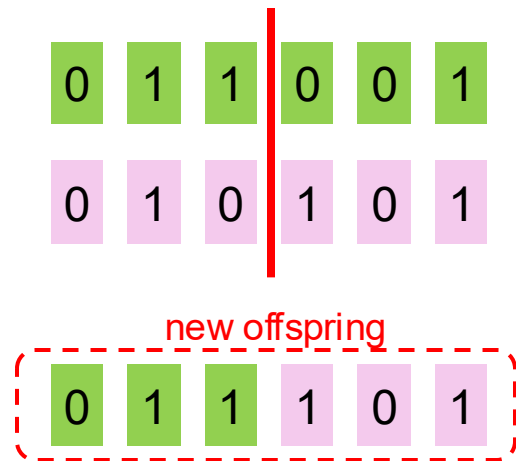
# Crossover (cont.)

## Definition

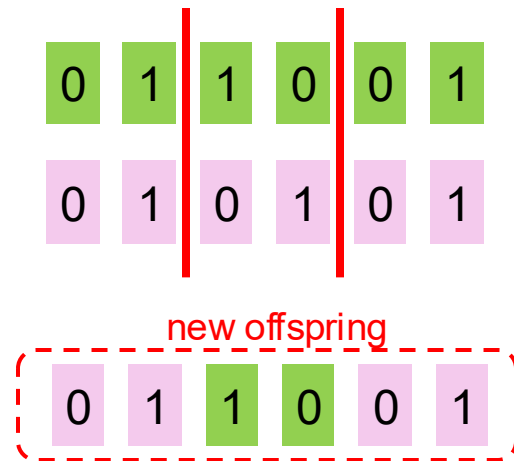
4

## Crossover

### single point



### Two point

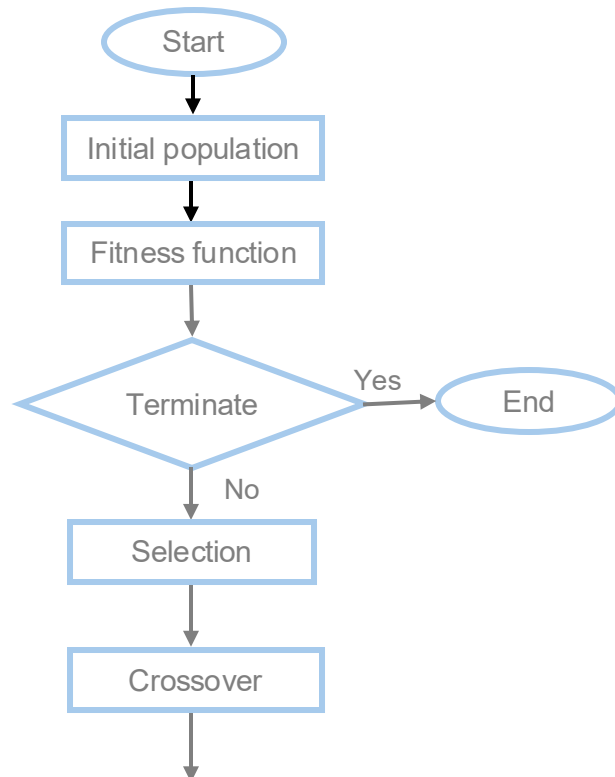


### Uniform



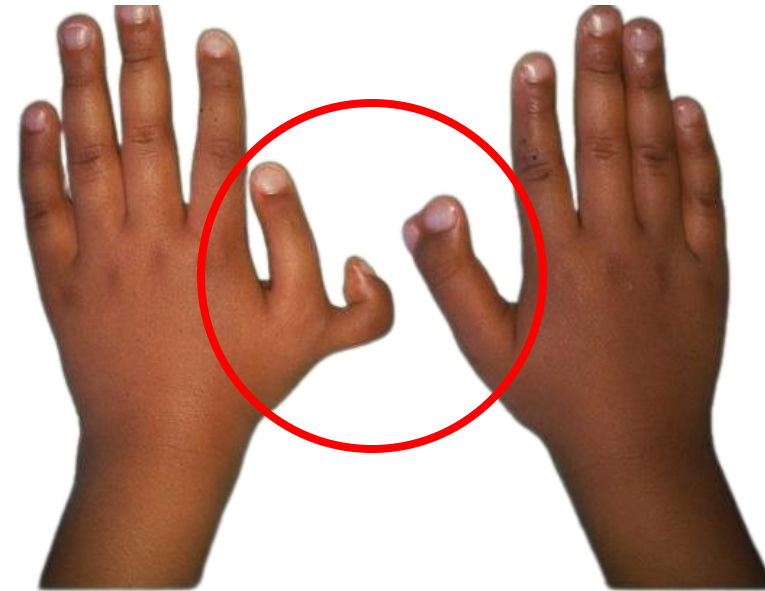
# Mutation

## Definition



**Mutation**

“**Mutation** is asexual reproduction to process the **generating new offspring with a single parent.**”



# Mutation (cont.)

## Definition

5

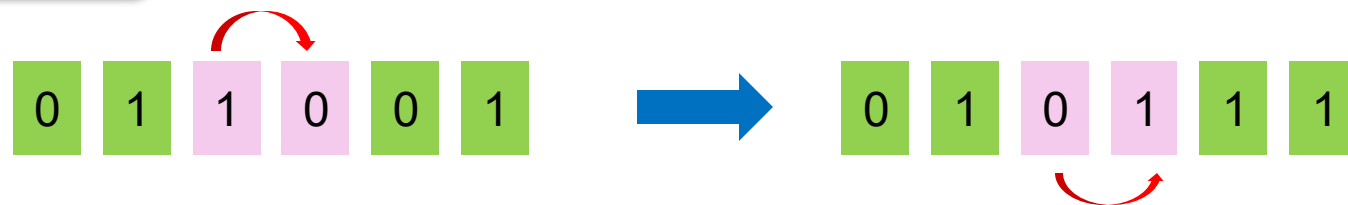
## Mutation

### Point

random point for mutation

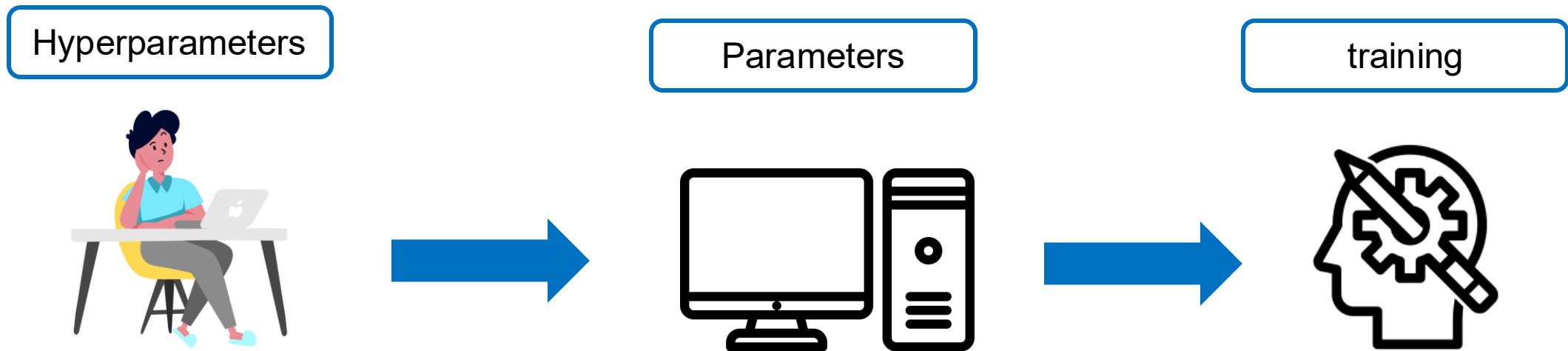


### Swap



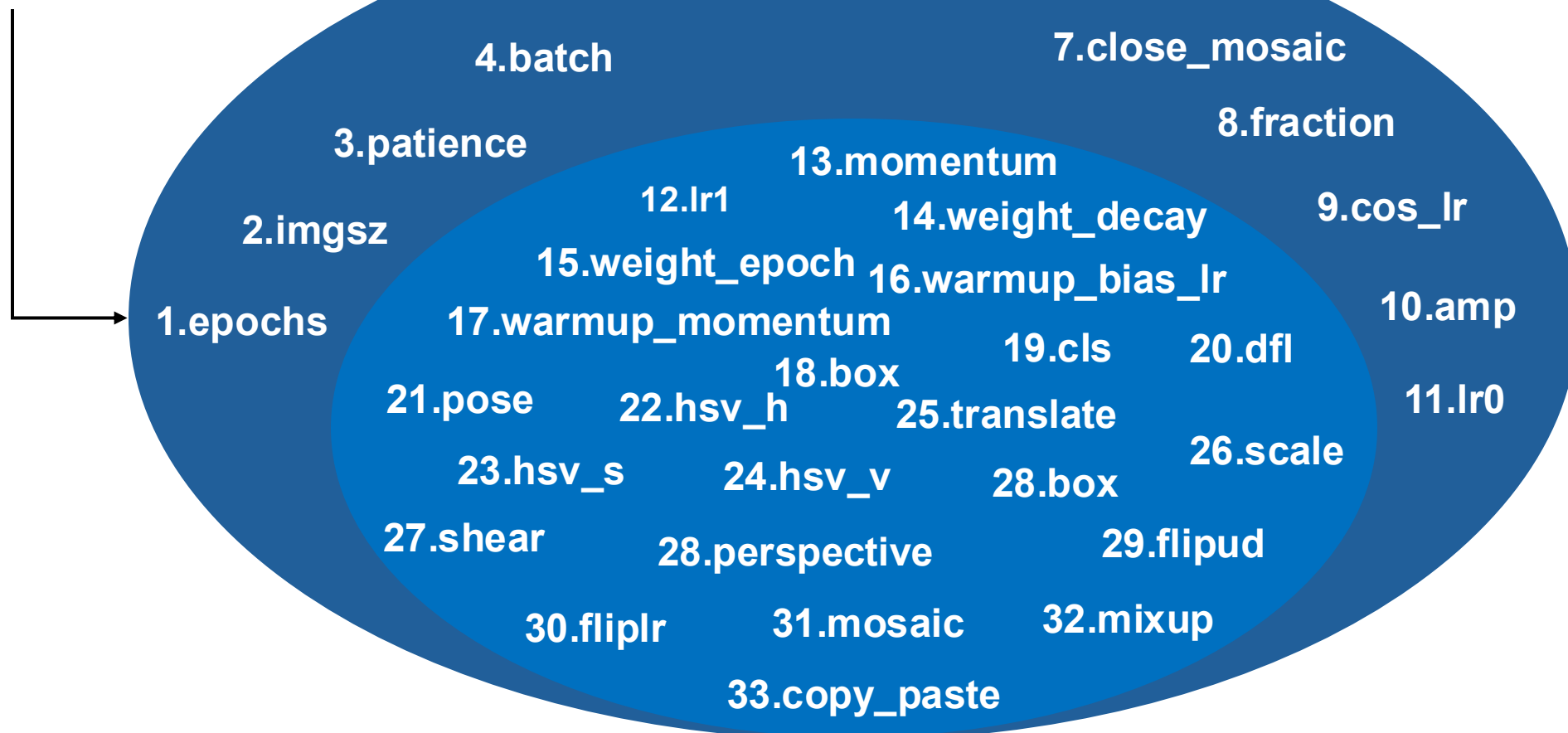
# Hyperparameter Tuning

“ **Hyperparameters** are parameters involve model building **before training process**”



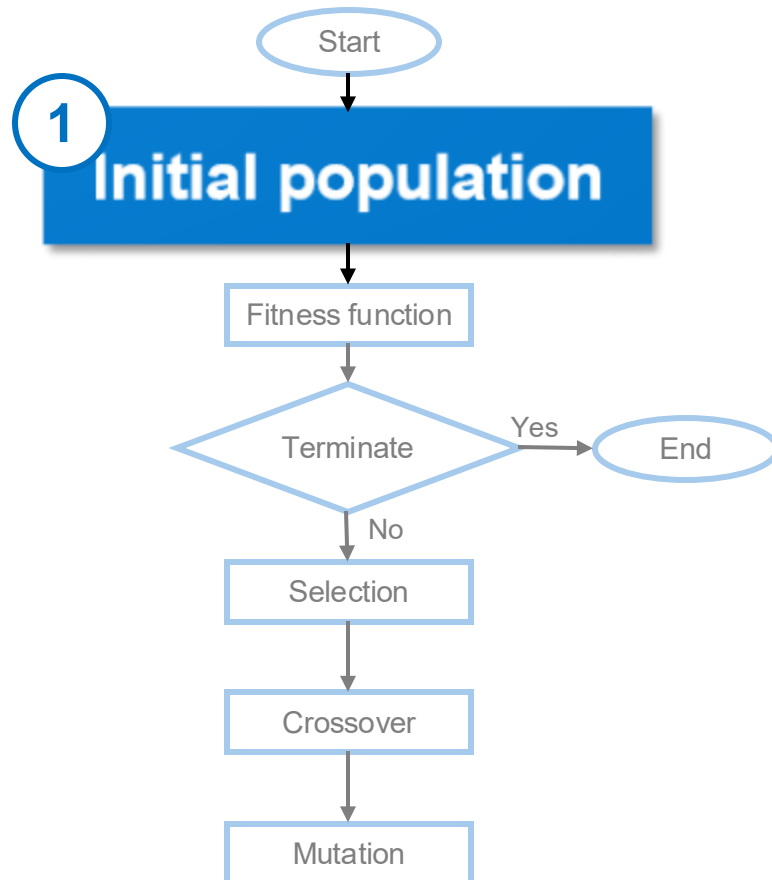
# Hyperparameter Tuning (cont.)

**All 33 hyperparameters**  
involved in YOLO.



# Example of Genetic Algorithm

## Definition

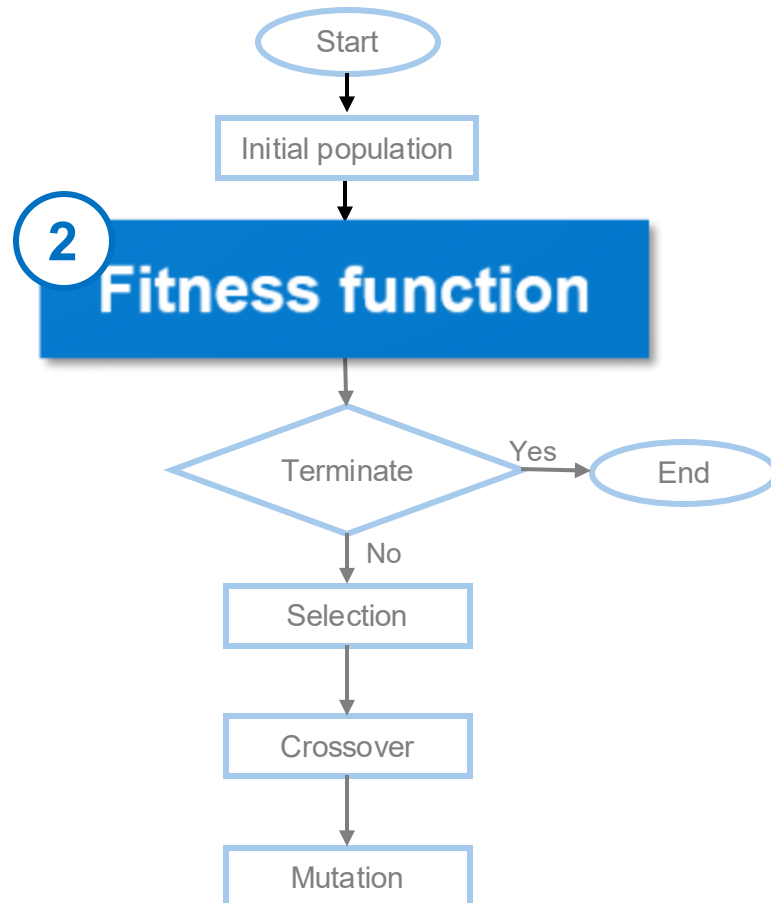


## Population

Parameter	Set 1	Set 2	Set 3
Initial learning rate	0.01	0.011	0.01132
Final learning rate	0.01	0.01	0.0048
Momentum	0.937	0.935	0.929
warmup epoch	3.0	2.838	2.578
Warmup momentum	0.8	0.787	0.436

# Example of Genetic Algorithm (cont.)

## Definition



In this example, we used “**Sensitivity**” to evaluate the performance.

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

### Evaluate Model

Use metrics like AP50, F1-score, or custom metrics to evaluate the model's performance.



# Example of Genetic Algorithm (cont.)

2

## Fitness function

Parameter	Initial learning rate	Final learning rate	Momentum	warmup epoch	Warmup momentum	Fitness value
Set 1	0.01	0.01	0.937	3.0	0.8	0.972
Set 2	0.011	0.01	0.935	2.835	0.787	0.969
Set 3	0.1132	0.0048	0.929	2.578	0.436	0.966

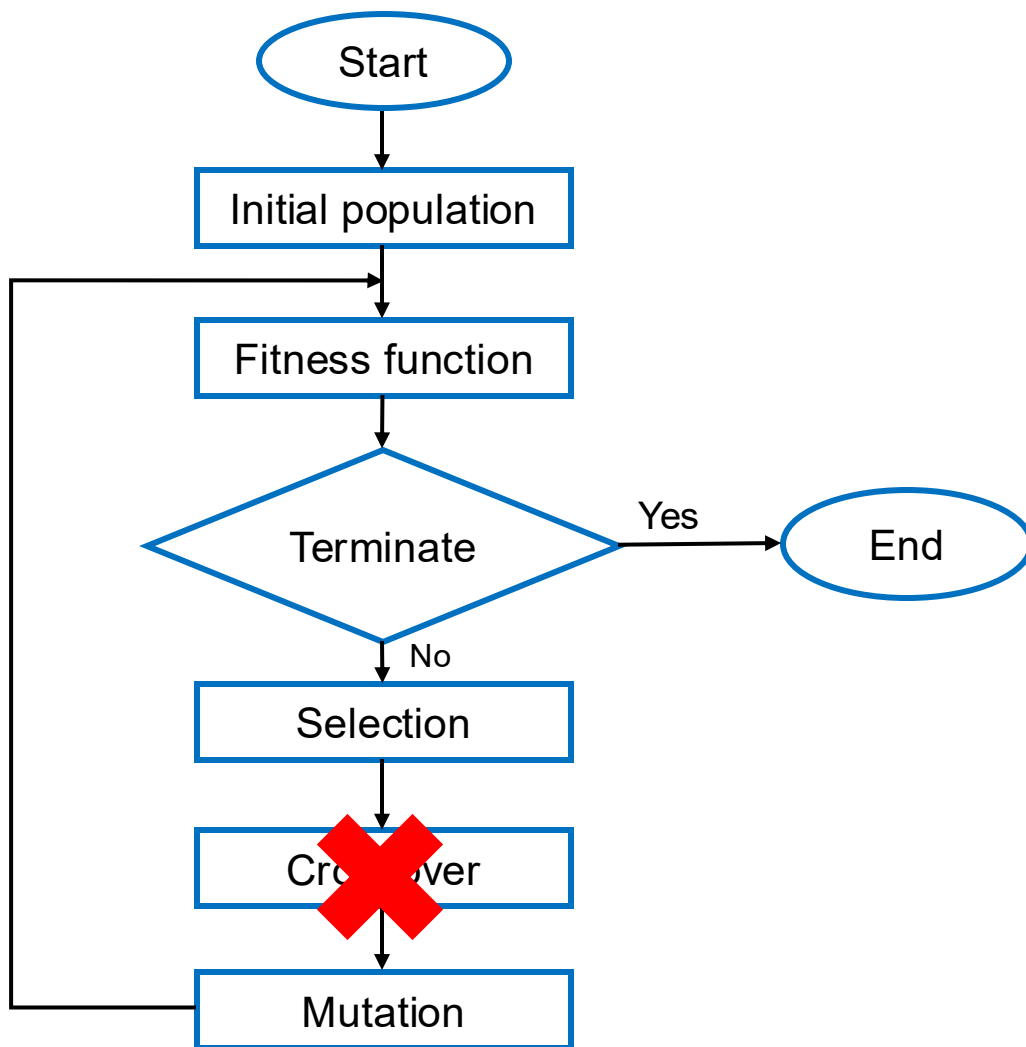
# Example of Genetic Algorithm (cont.)

3

## Selection

Parameter	P1	P2	P3	P4	P5	Fitness value
Set 1	0.01	0.01	0.937	3.0	0.8	<b>0.972</b>
Set 2	0.011	0.01	0.935	2.835	0.787	0.963
Set 3	0.1132	0.0048	0.929	2.578	0.436	<b>0.969</b>

# Remind that...



## Steps Involved

### Initialize Hyperparameters

Start with a reasonable set of initial hyperparameters. This could either be the default hyperparameters set by Ultralytics YOLO or something based on your domain knowledge or previous experiments.

### Mutate Hyperparameters

Use the `_mutate` method to produce a new set of hyperparameters based on the existing set.

### Train Model

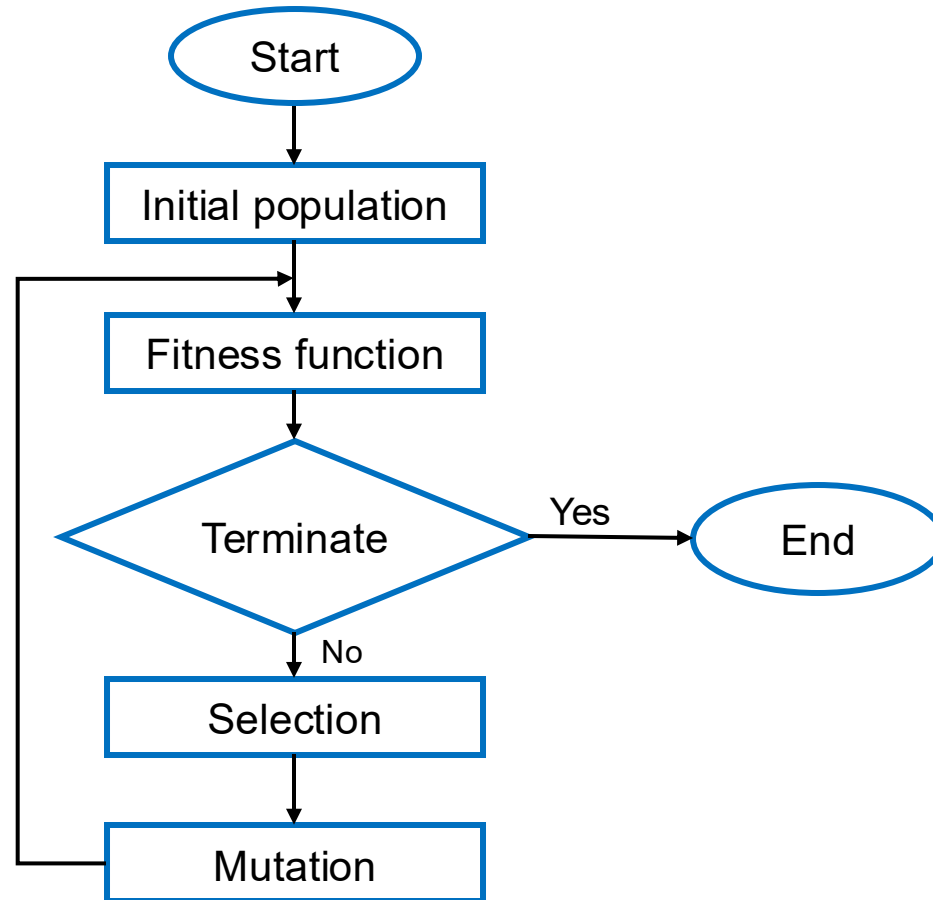
Training is performed using the mutated set of hyperparameters. The training performance is then assessed.

### Evaluate Model

Use metrics like AP50, F1-score, or custom metrics to evaluate the model's performance.

# Remind that...

## *Overall of genetic algorithm*



**Steps involve in hyperparameter based YOLO.**

# Example of Genetic Algorithm (cont.)

4

## Mutation

### Mutation based random point

Parameter	Initial learning rate	Final learning rate	Momentum	warmup epoch	Warmup momentum
Set 1	0.01	0.01	0.937	3.0	0.8
Set 3	0.1132	0.0048	0.929	2.578	0.436
Set 4 (new offspring)	0.01	0.01	0.978	2.760	0.5

# Example of Genetic Algorithm (cont.)

2

## Fitness function

Parameter	Initial learning rate	Final learning rate	Momentum	warmup epoch	Warmup momentum	Fitness value
Set 1	0.01	0.01	0.937	3.0	0.8	0.972
Set 3	0.1132	0.0048	0.929	2.578	0.436	0.966
Set 4	0.01	0.01	0.978	2.760	0.5	<b>0.980</b>

# Example of Genetic Algorithm (cont.)

3

Selection

**Reprocess...**

“If the **process isn't finished**, it will **continue iteration**”

Parameter	Initial learning rate	Final learning rate	Momentum	warmup epoch	Warmup momentum	Fitness value
Set 1	0.01	0.01	0.937	3.0	0.8	0.972
Set 3	0.1132	0.0048	0.929	2.578	0.436	0.966
<b>Set 4</b>	<b>0.01</b>	<b>0.01</b>	<b>0.978</b>	<b>2.760</b>	<b>0.5</b>	<b>0.980</b>

**Hyperparameter set 4** is optimal hyperparameter.

# Evaluation

		Actual	
		Fracture	No Fracture
Predicted	Fracture	TP	FP
	No fracture	FN	TN

The performance of the model is evaluated using the confusion matrix :

1. **True Positives (TP)** : The number of cases **correctly** predicted **as the target class**.
2. **True Negatives (TN)** : The number of cases **correctly** predicted **as not the target class**.
3. **False Positives (FP)** : The number of cases **incorrectly** predicted **as the target class**.
4. **False Negatives (FN)** : The number of cases **incorrectly** predicted **as not the target class**.

# Literatures Reviews

# AI in Hip Bone Diagnosis

- Automatic detection and functional subclassification of hip fractures using deep learning.

Category	Sensitivity
No fracture	0.937
Intertrochanteric fracture	0.923
FN fracture, displaced	0.896
FN fracture, non-displaced	0.512
Arthroplasty	0.979
ORIF	1.000

- Identification and classification of hip fracture.
- Based on Densely Connected Convolutional Neural Network (**DenseNet**).
- Trained on a dataset of **1,118 radiographs**
- Using data processing and augmentation.

The model achieves **an accuracy of 0.937** in hip fracture detection.

# AI in Hip Bone Diagnosis (cont.)

- Deep learning for femoral neck fracture detection on pelvic radiographs.

Parameter	Accuracy of 2 class	Accuracy of multiclass
Model	0.958	0.928
Expert 1	0.886	0.834
Expert 2	0.935	0.899
Expert 3	0.955	0.927
Resident 1	0.839	0.761
Resident 2	0.856	0.785
Resident 3	0.910	0.883

- **Femoral neck fracture detection** system based on a deep learning.
- Trained on a dataset of **1,494 radiographs**.
- Test on internal test set.

# AI in Hip Bone Diagnosis (cont.)

- Machine learning in classification of hip fracture.
  - Hip fracture classification.
  - Trained on a dataset of **3,659 radiographs**.
  - Based on Convolutional Neural Networks (CNNs) in **2 stage**.
    - **First stage** : localize the hip joint region
    - **Second stage** : using Grad-CAM to generate visual heatmap.

**The expert doctors**  
achieved **0.775 accuracy**.

**The model**  
achieved **0.920 accuracy**.

# AI in Hip Bone Diagnosis (cont.)

- Comparison of AI and Human Performance in Hip Fracture Detection.

**AI MODEL !**

**Sensitivity = 0.962**

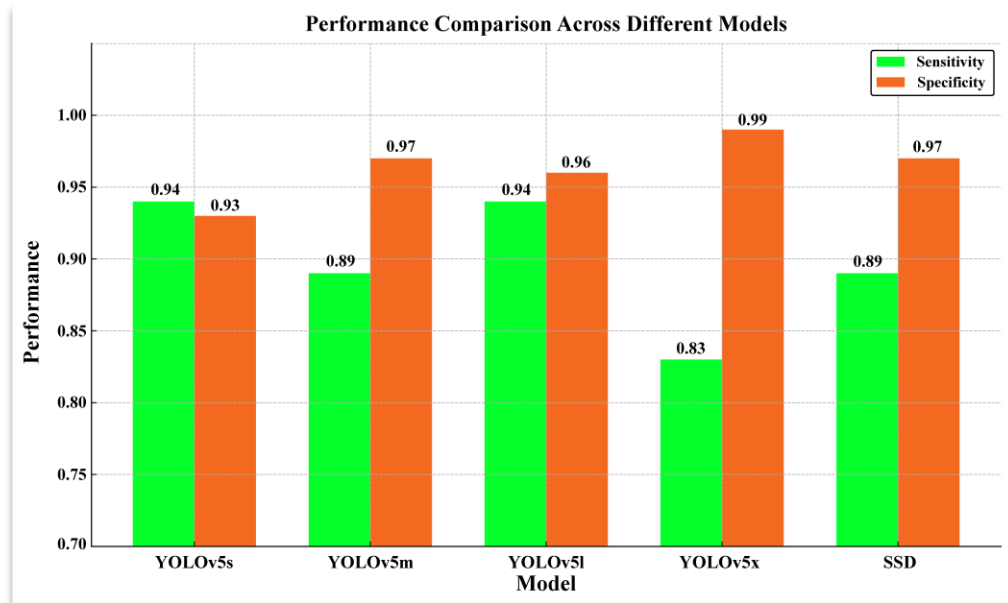
- Based on YOLOv4

human doctor	sensitivity
<b>Doctor, GP</b>	<b>0.622</b>
<b>Resident 1st, Ortho</b>	<b>0.711</b>
<b>Resident 1st, Radio</b>	<b>0.769</b>
Chief resident, Ortho	0.962
Chief resident, Radio	0.923
Attending, Ortho	0.962
Attending, Radio	0.962

"The model **performed similarly** to a well-trained **radiologist or orthopedist**"

# Improving Model Performance

➤ Deep learning model for bone fracture detection.



- Dysplasia of the hip (DDH) detection.
- Based on **YOLOv5** and **SSD algorithm**
- Trained on a dataset of **305 hip X-ray images**.
- Using data augmentation techniques.

**YOLOv5-L model** achieves **high performance** with a sensitivity of 0.94 and specificity of 0.96.

# Improving Model Performance (cont.)

- Non-displaced femoral neck fracture detection on hip radiographs.
  - Bone fracture classification
  - Based on **YOLOv8** and **VGG-16**.
  - Using **FracAtlas dataset**, which consists of 4,083 X-ray images.
  - Type of bone fractures, including **hand**, **leg**, **shoulder**, and **hip** fractures.

## YOLOv8

- After training for **50 epochs**, achieving an **0.810 training** accuracy and an **0.800 test accuracy**.

**VS**

## VGG-16

- **A maximum test accuracy of 0.730.**

# Improving Model Performance (cont.)

- Pediatric wrist trauma fracture detection based on deep learning.
  - Pediatric wrist trauma fracture detection.
  - **Comparing YOLO algorithm versions** with YOLOv5, YOLOv7, and YOLOv8.
  - Using a public dataset, specifically GRAZPEDWRI-DX, with **20,327 X-ray images** of wrist trauma in children.
  - Using data augmentation and data preprocessing.

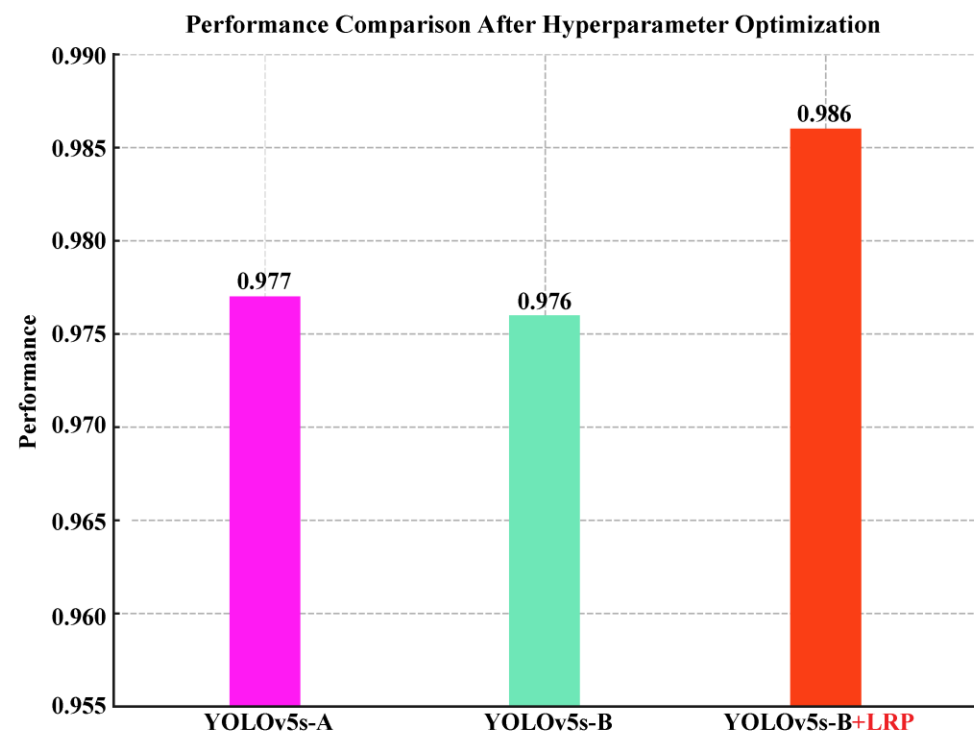
**YOLOv8** with a large model and **employ SGD optimization achieves mAP@50 of 0.638.**

# Hyperparameter Tuning

- Optimizing the hyperparameter tuning of YOLOv5 for underwater detection.

Hyperparameter configured based on **learning rate** and **momentum**

- the adaptive moment estimation (ADAM) optimizer
- the function reducing-learning-rate-on-plateau (LRP)



# Hyperparameter Tuning (cont.)

- Arecanut classification on X-ray images based on the YOLOv5 algorithm.
  - Arecanut classification based on the YOLOv5 algorithm.
  - Using hyperparameter tuning (13 hyperparameters)
  - Fine-tuned with GA and Adaptive Genetic Algorithm (AGA).

**The model with GA = mAP@50 of 0.956.**

**The model with AGA = mAP@50 of 0.977.**

\*AGA = GA with combines mutation and crossover

# Methodology

# Data Collection

## 1 Source of dataset :

- Collection of **1,811 hip bone X-ray images** from Srinagarind Hospital, Khon Kaen University.

## 2 Patient Demographics :

- Patients aged at least 18 years.
- Data collected between 2020 and 2022.

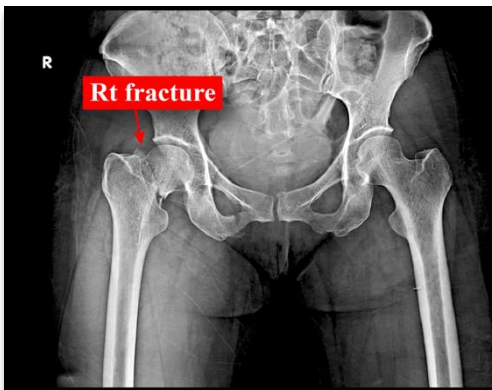
## 3 Data Privacy:

- All data are de-identified for patient information.
- **Approved by the KKU Ethics Committee to affirm** all methods adhered.

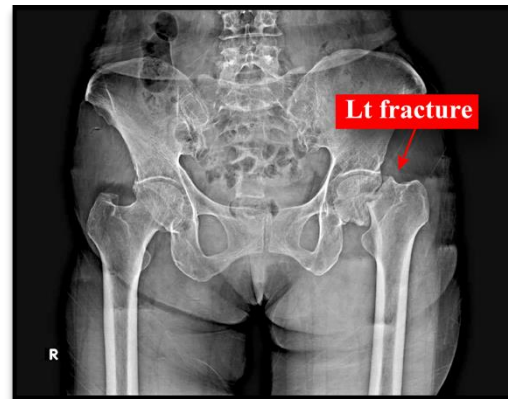
# Data Collection (cont.)

## Example of dataset

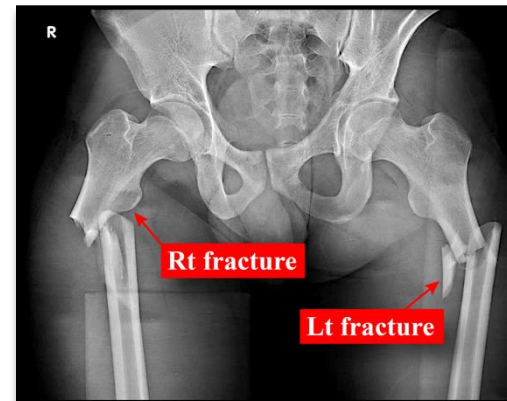
Fracture on  
right side



Fracture on  
left side



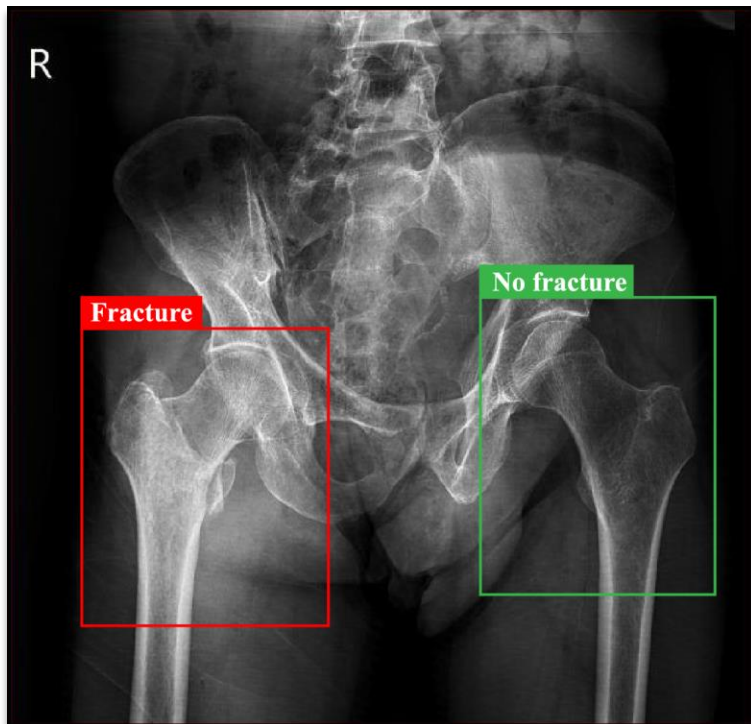
Fracture on  
both side



No fracture

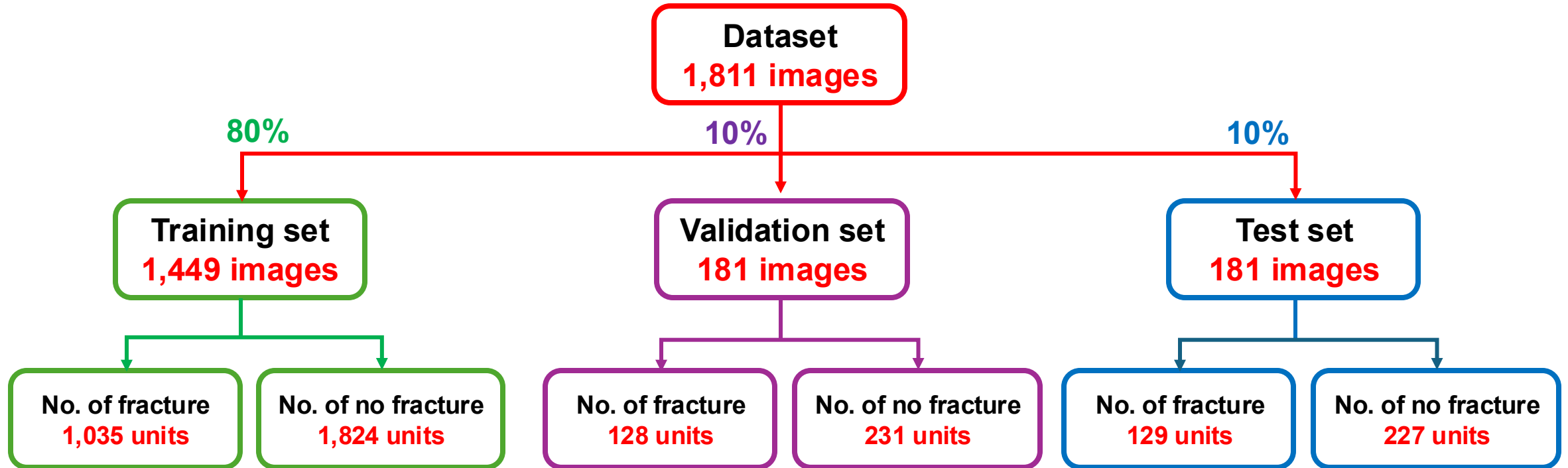


# Data Annotation

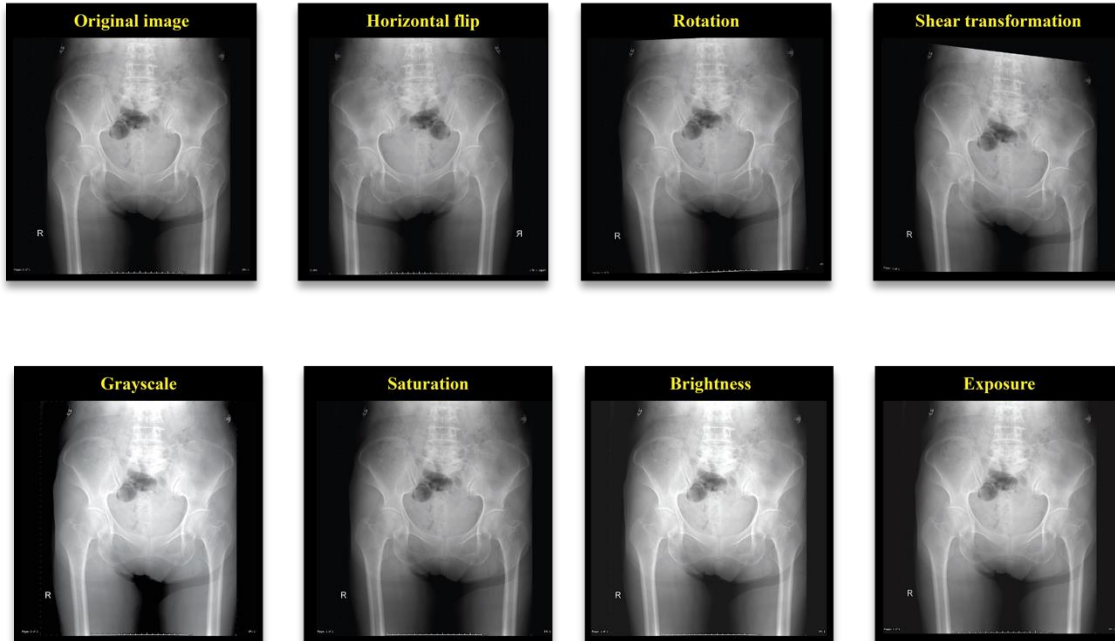


- All 1,811 hip X-ray images are manually annotated.
- Label types: **“Fracture”** and **“No fracture”**
- Annotation using Roboflow platform.

# Data Splitting



# Data Preprocessing



**\*\* The adjustment are based on real-world clinical condition.**

**The data preprocessing applied include:**

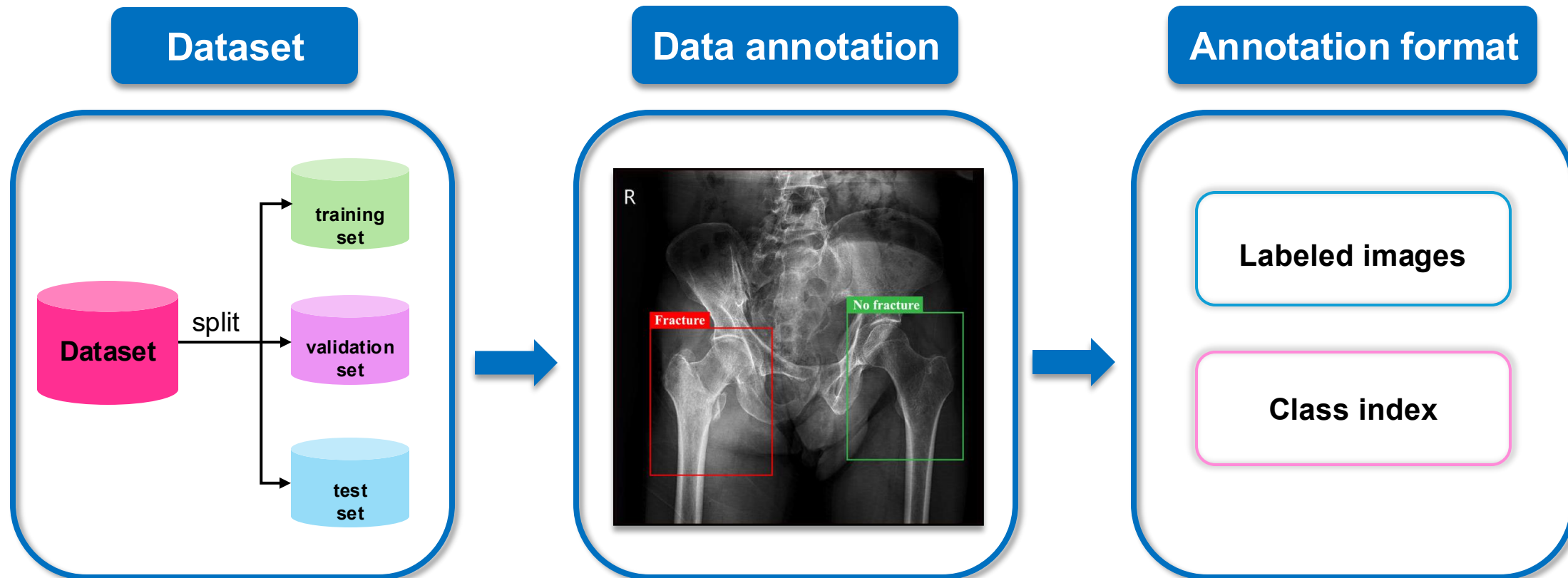
- **Horizontal flipping** :  $[-10^\circ, 10^\circ]$
- **Rotation** :  $[-10^\circ, 10^\circ]$
- **Shear transformation** :  $[-10^\circ, 10^\circ]$
- **Grayscale** : 25%
- **Saturation** : 30%
- **Brightness** : 10%
- **Exposure** : 10%

# Environment Setup

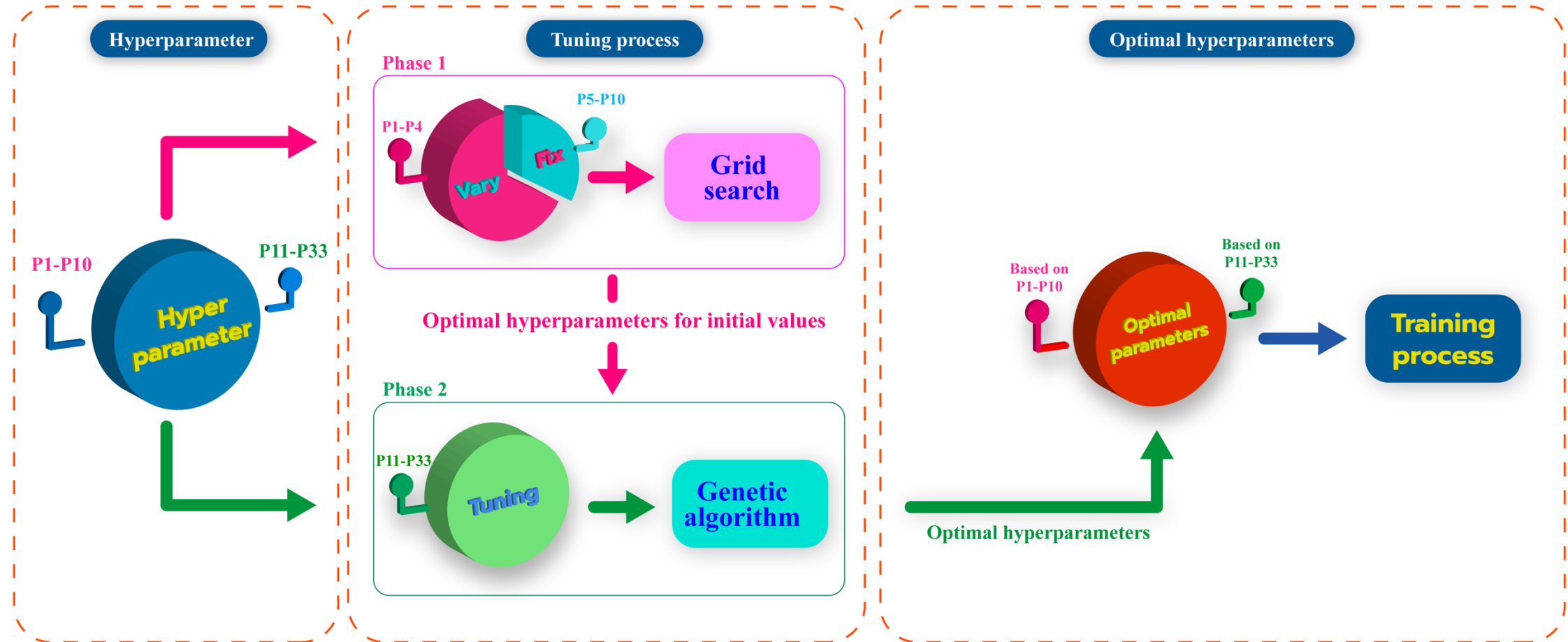
The table summarizes the hardware and software environments used for model training, including **local** and **cloud** computational.

Name	Local	Cloud
CPU	Intel Core i7-13700	Intel Xeon
GPU	NVIDIA RTX 4060	NVIDIA A100-SXM4
Programming Language	Python 3.12.9	Python 3.10.12
Programming Language	PyTorch 2.3.1	PyTorch 2.4.0
CUDA	12.4	12.1
Library Used	Ultralytics 8.1.27	Ultralytics 8.1.34

# Input Formatting



# Hyperparameter Tuning



# Hyperparameter Tuning (cont.)

***First stage : categorized into 2 groups.***

## Basic parameters

- Number of epochs
- Optimizer
- Batch size
- Image size
- Patience
- Seed
- AMP
- Fraction
- Close image mosaic epochs
- Cosine learning rate schedule

## Advanced parameters

- Initial learning rate
- Final learning rate
- Momentum
- Weight decay
- Warmup epoch
- Warmup momentum
- Box loss gain
- Class loss gain
- DFL loss gain
- HSV saturation augmentation
- HSV value augmentation
- HSV hue augmentation
- Image rotation
- Image translation
- Image scale
- Image shear
- Image perspective
- Image flip up-down
- Image flip left-right
- Image channel BGR
- Image mosaic
- Image mixup
- Copy-paste augmentation

# Hyperparameter Tuning (cont.)

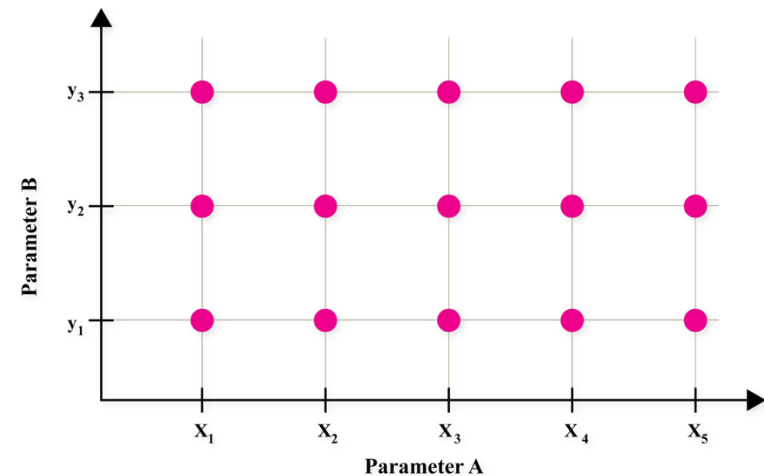
***Second stage : tuning process with grid search.***

## Basic parameters

- Number of epochs
- Optimizer
- Batch size
- Image size
- Patience
- Seed
- AMP
- Fraction
- Close image mosaic epochs
- Cosine learning rate schedule



## Tuning with grid search



# Hyperparameter Tuning (cont.)

***Second stage : tuning process with GA.***

## Advanced parameters

- Initial learning rate
- Final learning rate
- Momentum
- Weight decay
- Warmup epoch
- Warmup momentum
- Box loss gain
- Class loss gain
- DFL loss gain
- HSV saturation augmentation
- HSV value augmentation
- HSV hue augmentation
- Image rotation
- Image translation
- Image scale
- Image shear
- Image perspective
- Image flip up-down
- Image flip left-right
- Image channel BGR
- Image mosaic
- Image mixup
- Copy-paste augmentation



**Genetic  
algorithm**

# Hyperparameter Tuning (cont.)

## Third stage : Combination of Grid Search and GA

### Basic parameters

- Number of epochs
- Optimizer
- Batch size
- Image size
- Patience
- Seed
- AMP
- Fraction
- Close image mosaic epochs
- Cosine learning rate schedule

### Advanced parameters

- Initial learning rate
- Final learning rate
- Momentum
- Weight decay
- Warmup epoch
- Warmup momentum
- Box loss gain
- Class loss gain
- DFL loss gain
- HSV saturation augmentation
- HSV value augmentation
- HSV hue augmentation
- Image rotation
- Image translation
- Image scale
- Image shear
- Image perspective
- Image flip up-down
- Image flip left-right
- Image channel BGR
- Image mosaic
- Image mixup
- Copy-paste augmentation

**\*Optimal hyperparameters set for training process.**

# Evaluation

The test sets are collected from both internal and external medical institutions :

- **Internal datasets** : Srinagarind Hospital.
- **External datasets** : Chiang Mai Hospital, Wachira Hospital, and Songkla Hospital.

Test set	No. of images	No. of units	No. of fracture	No. of no fracture
Internal 1	14 images	28 units	14 units	14 units
Internal 2	181 images	356 units	129 units	227 units
Internal 3	378 images	710 units	375 units	335 units
External 1	31 images	62 units	31 units	31 units
External 2	40 images	78 units	40 units	38 units
External 3	500 images	1,000 units	500 units	500 units
<b>Total</b>	<b>1,144 images</b>	<b>2,234 units</b>	<b>1,089 units</b>	<b>1,145 units</b>

**\*\*All 1,144 hip X-ray images were ethically approved, with all patient information anonymized.**

# Results

# Model Performance

## *Model with hyperparameter set A*

Test set	No. of images	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Internal 1	14 images	0.757±0.0269	0.908±0.0920	0.586±0.0536	0.929±0.0781	0.706±0.0290
Internal 2	181 images	0.954±0.0082	0.950±0.0168	0.922±0.0237	0.973±0.0097	0.936±0.0121
Internal 3	378 images	0.959±0.0107	0.975±0.0057	0.946±0.0245	0.972±0.0067	0.960±0.0110
External 1	31 images	0.994±0.0128	0.994±0.0128	0.994±0.0128	0.994±0.0128	0.994±0.0128
External 2	40 images	0.915±0.0298	0.952±0.0358	0.879±0.0333	0.953±0.0349	0.914±0.0304
External 3	500 images	0.994±0.0036	0.997±0.0020	0.991±0.0060	0.997±0.0020	0.994±0.0036
<b>Total</b>	<b>1,144 images</b>	<b>0.970±0.0062</b>	<b>0.979±0.0023</b>	<b>0.956±0.0023</b>	<b>0.983±0.0154</b>	<b>0.967±0.0070</b>

# Model Performance (cont.)

## *Model with hyperparameter set B*

Test set	No. of images	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Internal 1	14 images	0.750±0.1052	0.929±0.0946	0.536±0.2062	0.965±0.0410	0.665±0.1790
Internal 2	181 images	0.968±0.0132	0.988±0.0049	0.922±0.0333	0.994±0.0029	0.954±0.0195
Internal 3	378 images	0.950±0.0122	0.977±0.0125	0.927±0.0159	0.975±0.0135	0.951±0.0115
External 1	31 images	0.996±0.0080	1.000±0.0000	0.992±0.0160	1.000±0.0000	0.996±0.0080
External 2	40 images	0.939±0.0289	0.951±0.0566	0.931±0.0237	0.947±0.0615	0.940±0.0265
External 3	500 images	0.996±0.0023	0.999±0.0010	0.993±0.0048	0.999±0.0010	0.996±0.0024
<b>Total</b>	<b>1,144 images</b>	<b>0.971±0.0074</b>	<b>0.988±0.0061</b>	<b>0.952±0.0111</b>	<b>0.988±0.0057</b>	<b>0.968±0.0082</b>

# Model Performance (cont.)

## *Model with hyperparameter set C*

Test set	No. of images	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Internal 1	14 images	0.743±0.0269	0.928±0.0592	0.528±0.0348	0.957±0.0348	0.672±0.0353
Internal 2	181 images	0.948±0.0060	0.947±0.0184	0.908±0.0153	0.971±0.0101	0.926±0.0085
Internal 3	378 images	0.899±0.0107	0.898±0.0252	0.913±0.0222	0.883±0.0330	0.905±0.0095
External 1	31 images	0.974±0.0128	0.969±0.0196	0.981±0.0157	0.968±0.0206	0.974±0.0128
External 2	40 images	0.890±0.0114	0.933±0.0396	0.853±0.0510	0.930±0.0438	0.889±0.0145
External 3	500 images	0.983±0.0047	0.980±0.0085	0.986±0.0068	0.980±0.0087	0.983±0.0047
<b>Total</b>	<b>1,144 images</b>	<b>0.943±0.0055</b>	<b>0.945±0.1571</b>	<b>0.939±0.1120</b>	<b>0.944±0.0171</b>	<b>0.941±0.0052</b>

# Model Performance (cont.)

## *Model with hyperparameter set D*

Test set	No. of images	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Internal 1	14 images	0.973±0.0265	0.980±0.0040	0.600±0.0573	0.986±0.0284	0.742±0.0431
Internal 2	181 images	0.974±0.0037	0.966±0.0098	0.961±0.0072	0.980±0.0060	0.964±0.0052
Internal 3	378 images	0.951±0.0124	0.973±0.0074	0.935±0.0222	0.971±0.0081	0.953±0.0127
External 1	31 images	0.987±0.0123	0.993±0.0132	0.981±0.0158	0.994±0.0128	0.987±0.0123
External 2	40 images	0.963±0.0214	1.000±0.0000	0.928±0.0416	1.000±0.0000	0.962±0.0222
External 3	500 images	0.997±0.0025	0.998±0.0020	0.996±0.0045	0.998±0.0020	0.997±0.0025
<b>Total</b>	<b>1,144 images</b>	<b>0.974±0.0064</b>	<b>0.984±0.0041</b>	<b>0.962±0.0114</b>	<b>0.986±0.0035</b>	<b>0.973±0.0067</b>

# Model Performance (cont.)

## *Model with hyperparameter set E*

Test set	No. of images	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Internal 1	14 images	0.750±0.0503	0.946±0.0659	0.529±0.0727	0.972±0.0348	0.677±0.0731
Internal 2	181 images	0.957±0.0097	0.975±0.0122	0.905±0.0319	0.987±0.0070	0.938±0.0158
Internal 3	378 images	0.928±0.0138	0.957±0.0178	0.906±0.0184	0.954±0.0190	0.930±0.0138
External 1	31 images	0.984±0.0179	0.981±0.0155	0.987±0.0260	0.981±0.0158	0.984±0.0179
External 2	40 images	0.895±0.0169	0.939±0.0290	0.849±0.0254	0.942±0.0303	0.891±0.0169
External 3	500 images	0.982±0.0126	0.988±0.0084	0.976±0.0223	0.988±0.0083	0.982±0.0130
<b>Total</b>	<b>1,144 images</b>	<b>0.952±0.0121</b>	<b>0.971±0.0067</b>	<b>0.929±0.0238</b>	<b>0.971±0.0065</b>	<b>0.948±0.0139</b>

# Model Performance (cont.)

## *Model with hyperparameter set F*

Test set	No. of images	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Internal 1	14 images	0.807±0.0171	0.942±0.0476	0.657±0.0284	0.957±0.0348	0.773±0.0199
Internal 2	181 images	0.970±0.0044	0.960±0.0204	0.956±0.0135	0.977±0.0122	0.958±0.0054
Internal 3	378 images	0.966±0.0021	0.977±0.0096	0.960±0.0123	0.974±0.0108	0.968±0.0022
External 1	31 images	1.000±0.0000	1.000±0.0000	1.000±0.0000	1.000±0.0000	1.000±0.0000
External 2	40 images	0.974±0.0160	0.980±0.0184	0.970±0.0245	0.979±0.0198	0.975±0.0160
External 3	500 images	0.996±0.0006	0.996±0.0020	0.996±0.0020	0.996±0.0020	0.996±0.0006
<b>Total</b>	<b>1,144 images</b>	<b>0.979±0.0012</b>	<b>0.983±0.0063</b>	<b>0.973±0.0062</b>	<b>0.985±0.0056</b>	<b>0.977±0.0014</b>

# Model Performance (cont.)

## *Model with hyperparameter set G*

Test set	No. of images	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Internal 1	14 images	0.829±0.0348	0.980±0.0400	0.671±0.0572	0.986±0.0284	0.795±0.0469
Internal 2	181 images	0.978±0.0004	0.979±0.0064	0.958±0.0064	0.989±0.0036	0.969±0.0004
Internal 3	378 images	0.969±0.0055	0.981±0.0062	0.959±0.0052	0.980±0.0067	0.970±0.0051
External 1	31 images	0.997±0.0064	0.994±0.0124	1.000±0.0000	0.994±0.0128	0.997±0.0064
External 2	40 images	0.954±0.0102	0.984±0.0127	0.924±0.0149	0.984±0.0127	0.953±0.0098
External 3	500 images	0.998±0.0010	0.998±0.0020	0.998±0.0000	0.998±0.0020	0.998±0.0010
<b>Total</b>	<b>1,144 images</b>	<b>0.982±0.0017</b>	<b>0.989±0.0014</b>	<b>0.972±0.0027</b>	<b>0.989±0.0012</b>	<b>0.980±0.0016</b>

**Optimal hyperparameters set for AI-based hip fracture detection.**

# Model Performance (cont.)

## *Model with hyperparameter set H*

Test set	No. of images	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Internal 1	14 images	0.757±0.0269	0.975±0.0500	0.528±0.0348	0.986±0.0284	0.685±0.0363
Internal 2	181 images	0.971±0.0087	0.973±0.0139	0.947±0.0124	0.985±0.0075	0.960±0.0125
Internal 3	378 images	0.960±0.0066	0.991±0.0055	0.933±0.0163	0.991±0.0063	0.961±0.0065
External 1	31 images	0.997±0.0064	1.000±0.0000	0.994±0.0128	1.000±0.0000	0.997±0.0064
External 2	40 images	0.923±0.0269	0.963±0.0263	0.885±0.0515	0.963±0.0269	0.921±0.0283
External 3	500 images	0.998±0.0008	0.999±0.0016	0.998±0.0000	0.999±0.0016	0.998±0.0008
<b>Total</b>	<b>1,144 images</b>	<b>0.976±0.0030</b>	<b>0.991±0.0029</b>	<b>0.958±0.0067</b>	<b>0.993±0.0025</b>	<b>0.973±0.0034</b>

# Limitations

## 1 Limit of dataset

- The dataset for model training includes only the anteroposterior (AP) pelvis view.
- The models do not support other clinical perspectives, such as the lateral trans-inguinal view.

## 2 Limit of computational resource

- The model training process requires high computational resources and time-consuming.
- The fastest training case is approximately 8 hours and complex model extended the training time up to 1 month.
- Limitation factor for wider experimentation.

## 3 Limit of computational resource

- The model not support X-ray images captured through non-clinical data, such as smartphone cameras

# Limitations and Future Works

# Limitations

## 1 Limit of dataset

- The dataset for model training includes only the anteroposterior (AP) pelvis view.
- The models do not support other clinical perspectives, such as the lateral trans-inguinal view.

## 2 Limit of computational resource

- The model training process requires high computational resources and time-consuming.
- The fastest training case is approximately 8 hours and complex model extended the training time up to 1 month.
- Limitation factor for wider experimentation.

## 3 Limit of computational resource

- The model not support X-ray images captured through non-clinical data, such as smartphone cameras

# Future Works

- 1** **To develop the model** with image processing and **apply visual explanation with a heatmap** to assist human doctors in decision-making.
- 2** **To explore alternative optimization techniques** that can improve both model performance and computational efficiency in hyperparameter tuning.